U.PORTO

FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

EXPLORATORY ANALYSIS: THE IMPORTANCE OF THE EGO IN A TRUST NETWORK

By

Marta Pinto Leite Maximiano Ferreira

Dissertation of Master in Data Analytics

Supervised by:

Professor João Manuel Portela Gama

Fabíola Souza Fernandes Pereira

Faculdade de Economia

Universidade do Porto

2017

Biographical note

Marta Pinto Leite Maximiano Ferreira, born on November 3rd, 1992, in Porto, completed her degree in Management at Faculty of Economics of University of Porto (FEP.UP), in July 2014.

Motivated by the desire to learn and explore more, especially in what relates to data analysis, in September 2015 joined the Master in Data Analytics at Faculty of Economics of University of Porto.

Acknowledgments

I would like to begin by thanking professor João Gama and Fabíola for their support and suggestions throughout this year.

To my parents and family, who always supported and motivated me.

To Marco, for all the love, support and tolerance, and Patrícia, who also accompanied me throughout this journey, for the friendship and patience.

Abstract

Nowadays, the use of Internet to gather information about certain goods or services is part of the routine of almost everyone. Before choosing the restaurant that best suits someone interests, anyone may quickly search online for recommendations on multiple restaurants, or, before buying a product they can look for other peoples' reviews about it and decide if it is worth buying.

This search for other users' opinions has led to the emergence of several reviews websites.

This work will be based on the analysis of a network extracted from the website Epinions.com, which was an online product rating website where people would give indication of which users they trusted the most, based on their ratings and reviews, creating their own "web of trust" network. The formation of these networks allows us now to analyze a trust network in the field of online social networks.

What we aim to analyze is, from an egocentric perspective, the impact that the presence or absence of the Ego has on the other nodes of the network, its structure and metrics.

Keywords: **s**ocial networks; ego networks; trust networks; Epinions; communities.

Resumo

Nos dias de hoje, recorrer à Internet para obter informações sobre determinados bens ou serviços já faz parte da rotina de quase toda a gente. Por exemplo, antes de escolhermos onde vamos jantar, podemos procurar na Internet as opiniões de outras pessoas sobre certo restaurante, ou ainda antes de comprarmos algum produto podemos procurar saber o que é que as outras pessoas que já adquiriram esse produto pensam e se valerá a pena comprá-lo, conforme as opiniões *online*.

Esta busca pela opinião das outras pessoas levou à criação de vários *websites* de opinião.

Este trabalho vai ser baseado na análise de uma rede extraída do *website* Epinions.com, que era um *website* de classificação de produtos, onde as pessoas podiam indicar quais os utilizadores em cujas opiniões mais confiavam, conforme as classificações e comentários a certos produtos, criando, assim, uma rede de confiança. A formação destas redes permite-nos, agora, analisar uma rede de confiança no âmbito da análise de redes sociais.

O que pretendemos analisar é, de uma perspetiva egocêntrica, qual o impacto que a presença ou ausência do Ego tem nos outros nós da rede, na sua estrutura e métricas.

Palavras-chave: redes sociais; ego redes; redes de confiança; Epinions; comunidades.

Table of Contents

List of Figures

List of Tables

1 Introduction

A Social Network consists in a set of actors and the relations they establish between them, which do not have to be necessarily social or on-line interactions; they may represent information flows, transfers of goods, cooperation between individuals, among others.

Social Network Analysis combines social sciences with network analysis and, as said by Koçak, NG (2014), a large part of the terminology used in it has been drawn or adapted from graph theory and has two main approaches: the sociocentric and egocentric approach, which will be the focus of this dissertation.

Through Social Network Analysis, which, as mentioned, has two approaches, the *egocentric* – which studies the relations surrounding individuals rather than the whole society – and the *sociocentric* – which focuses on the whole network and their relationships (Davis, S. Chung, & Hossain, 2006), by analyzing the Ego Network of specific users individually, the goal is to analyze the impact of the Ego's presence/absence on its own trust network.

1.1 Motivation

Nowadays, the use of Internet to gather information about certain goods or services is part of the routine of almost everyone. Before choosing the restaurant that best suits someone interests, anyone may quickly search online for recommendations on multiple restaurants, or, before buying a product they can look for other peoples' reviews about it and decide if it is worth buying.

This search for other users' opinions has led to the emergence of several reviews websites.

This work will be based on the analysis of a network extracted from the website *Epinions.com*, which was an online product rating website where people could perceive the opinion of the other users about a particular good and, additionally, they were able to indicate which users they trusted the most, based on their ratings and reviews, creating their own “web of trust”. The formation of these networks allows us now to analyze a trust network in the field of online social networks.

The motivation for the development of this dissertation arose from the desire to apply the knowledge acquired during the master’s classes to a real case, essentially because not only social network analysis was a theme that drew my attention from the beginning but it is also such a current theme yet with a lot to explore.

1.2 Structure

This work has the following structure: the first chapter is an introduction to the topic we want to analyze; the second chapter is divided into three sub-topics: the first one is about social network analysis and explores the main metrics used to analyze social networks, both in a sociocentric and egocentric perspective, and also gives us an highlight to some of the main concepts of Social Network Analysis, so that it is possible to understand the whole content of this work; the second one explores the concept of “trust” and the way people trust in online social networks; and the third one gives some examples of studies already published about these topics; the third chapter is our practical application, where we analyze the network in several ways, extracting from it as much information as we can and think it’s useful; and the fourth chapter presents our conclusions.

2 Trust and Social Networks Analysis

2.1 Social Networks Analysis

In 1932, at the New York Training School for Girls, a reformatory school for teenage girls, within two weeks, 14 girls ran away from the institution. In order to study why this had happened, Moreno (1934) used the sociograms method to study 500 girls. He studied their behavior and feelings towards each other and was able to create a social map, where he could see the relationships that existed between the girls. He concluded that their behavior was due to the position occupied in the sociogram.

This study proves how the analysis of social maps – social networks – can tell us a lot about the behaviors of the entities whose networks we are analyzing.

In this chapter, there will be highlighted some key concepts, as well as some studies carried out in the scope of Ego Networks as well as their metrics and detection of communities.

2.1.1 **Key concepts**

For the content of this work to be perceptible, there are some concepts about Social Networks that should be emphasized.

As previously mentioned, the graphs theory is one of the pillars of Social Network Analysis. A graph is a set of vertices and edges, where each edge represents a connection between two vertices and each vertex represents a social entity.

There are two main structures for the storage and subsequent representation of social networks, according to our data and what we want to analyze, which are list structures and matrix structures. Matrix structures are most used when dealing with dense matrices since they consume a lot of storage space and, for instance, assuming a graph of 10 vertices but where only 2 of them are interconnected, the matrix would be composed mostly of zeros but still the storage space consumed would be the same as if all the vertices were interconnected. List structures are specially used when dealing with sparse graphs since they consume much less storage space. (Oliveira & Gama, 2012) **Figure 1** is an example of the representation of a social network in a graph (a)), an adjacency list (b)) and a matrix (c)).

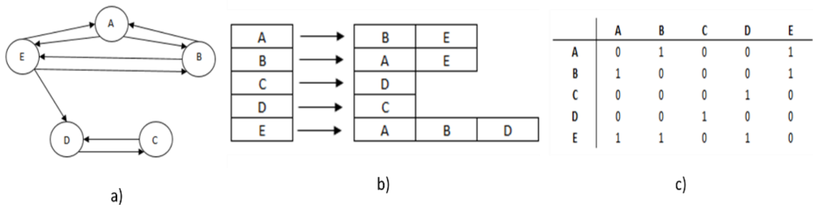

Figure 1 - Social network data different representations

Graphs may be classified according to their direction, this is, they might be *direct* if each edge represents the direction of influence of a social entity, or *undirected*, if the edges represent bilateral connections; they may even be *cyclic*, if there is a path that starts and ends at a given node, or *acyclic,* otherwise.

Vertices are usually called as actors, being an actor a social entity that is part of a network and establishes relationships with other social entities; they might represent individuals, entities, objects or organizations. The relationship between actors is represented by edges, which characterize, thus, a relational tie between actors.

It is called a dyad or triad when a network is formed by two vertices – or three, if we are talking about triads – that may or may not be connected to each other. When none of the vertices is connected, it is called a null dyad; when a connection is unilateral, it is called an asymmetric dyad, and when the connection is bilateral, it is a mutual dyad. A dyad is "the smallest social structure in which an individual can be embedded". (Hanneman & Riddle, 2005)

A group is a set of all vertices of a network and their relationships and a subgroup is a subset of a group.

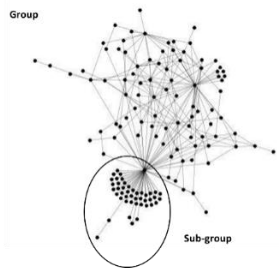

Figure 2 - Example of a group and sub-group

When, in a subgroup, each actor is connected to all the other actors at a geodesic distance equal to 1, it is said to be a clique. There is also the concept of N-clique, which is when each actor is connected to all the others within a distance of N.

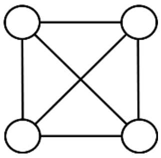

Figure 3 - Example of a clique

Figure 4 is an example of an indirect network in which not all nodes are connected. This network has three subgraphs where the nodes are only connected among themselves and are not connected to any other node from the other subgraphs. When this happens, we call these subgraphs *connected components*.

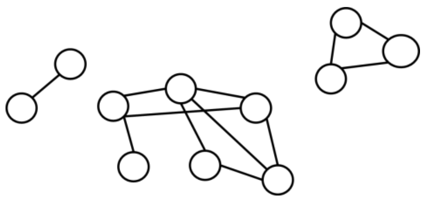

Figure 4 - Example of connected components – indirect networks

In **Figure 5** we have an example of a direct network and we can see that not every node is reachable or reaches all the other nodes. As we can see, nodes 4 and 5 are mutually connected and do not connect to any other node – node 4 receives a link from node 3 but does not link back –, this is, node 4 is connected to node 5 and node 5 is connected to node 4 – when this happens, it is called a *strongly connected component*; the same way, nodes 1, 6 and 3 also form a strongly connected component, since node 3 connects to node 1 and 6, node 1 connects to node 6 and node 6 connects to node 3.

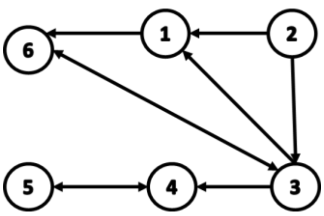

Figure 5 - Example of connected components - direct network

2.1.2 **Ego Networks**

An Ego Network consists of a central actor, whose network we are analyzing, his friends and the friends of his friends, according to the depth we can gather from the network, which can be of 1, 1.5 or 2. If the network has depth 1, the network only shows the direct connections between the focal actor and his friends; if it has depth of 1.5, it means the network shows the direct connections between the focal actor and his friends as well as the connections between the friends themselves; lastly, if it has depth 2, it means the network shows the connection of depth 1,5 plus the connections that his friends have with other people, i.e., in the network will also appear his friends' friends. (Hirst, 2010)

In the egocentric approach, we refer to the focal actor as the "Ego" and the nodes to whom the Ego is directly connected are called "Alters". (Halgin & Borgatti, 2012) So, in other words, an Ego Network is formed by the Ego, the alters and the alters of the alters.

The actors that are positioned further away from the Ego are his "acquaintances" while those who are closer to him are his "friends". Usually, the Ego's friends are a part of a "close-knit" group, and often everyone knows each other. The acquaintances are less likely to know one another.

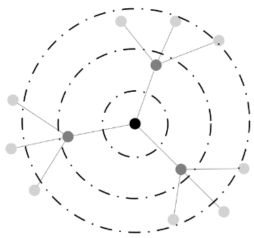

Figure 6 – Example of Ego Network
(From: https://www.thinglink.com/scene/479486734900396033)

In **Figure 6**, we can see an Ego Network divided into three "dimensions": in the first dimension, in the center, we have the Ego; in the second one, we have three nodes, who represent the friends of the Ego; and in the third one, more distant, we have the friends of the friends, who are the acquaintances of the Ego.

The existence of this type of connections, closer or further away from the Ego, leads us to the concept of Core-Periphery Network. The group that is most interconnected and closer to the Ego is called *Core*, this is where the Ego and his friends are, and the group that is more separated is called *Periphery*, which is where the acquaintances of the Ego are located.

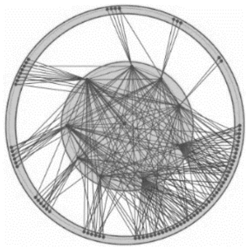

Figure 7 – Example of Core-Periphery structure
(From: http://rf.mokslasplius.lt/en/core-periphery-network-models)

Starting from the study of Ego Networks, there are some properties of global Social Networks that can be derived, as we will see ahead. For this, we use some metrics that will be presented in the next topic.

2.1.3 **Metrics**

2.1.3.1 Global Networks

To improve Social Network Analysis and understand its behavior, it's important to study its statistical measures. This analysis becomes very useful since it allows to obtain information about the network without its graphical representation.

They will be divided between measures of centrality of the actors – an analysis at the node level – and the measures of the network structure. But first there are some concepts that should be highlighted:

- The *Weak Ties Theory* is a common feature to all Social Networks and states that there may be an actor that makes an interconnection between two different "close-knit" groups. Granovetter (1983) defends that "the weak tie between Ego and his acquaintance becomes not merely a trivial acquaintance tie but rather a crucial bridge between the two densely knit clumps of friends", since it explains the transmission of information between groups.

The connection formed between two actors of different groups is called “bridge” or “gatekeeper”. A “bridge” is a connection between two different groups, whose intermediaries are usually known as “brokers” and act as bridges in structural holes, this is, in gaps between two actors of different groups.

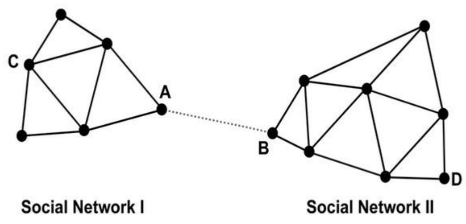

Figure 8 – Example of Weak Ties Theory

In **Figure 8**, we can see that actors A and B act as bridges between Social Network I and Social Network II and if it wasn’t for their connection, there would be a structural gap between them;

- The principle of *Transitivity* states that in a triad, for instance, if actor A is linked with actors B and C, and B is only linked with A, then, there is a high chance of actors B and C establish a connection.

  The measure for transitivity is defined by:

$$Transitivity\ index = \frac{No.of\ transitive\ triads}{No.of\ potentially\ transitive\ triads}$$

  Where,

  - *Transitive triads* are triads whose actors are all connected
  - *Potentially transitive* are triads where two of the actors have the possibility to establish a connection (Snijders, 2012)

  And its result varies between 0 and 1. It is 0 when there are none triads formed and 1 when all triads are complete;

  Transitivity is an evidence for the existence of strong ties, but not a necessary or sufficient condition.

- *Homophily* is the tendency of an actor to establish a relationship with other actor whose characteristics and preferences are quite similar. This causes homogeneous groups – *clusters* – to form.

  There are two reasons that motivate actors to connect with other actors:

    - Selection – “people tend to form friendships with others who are like them”, whether because they are the same age, same sex, etc. (Easley & Kleinberg, 2010)
    - Social Influence – people may change some of their characteristics, such as opinions and behaviors, in order to become more similar to their friends.

  Homophily and Transitivity together lead to the formation of cliques – clusters that are fully connected – since homophily leads to the formation of clusters and, therefore, there is a higher chance that every actor connects with all the others.

- A *path* is a sequence of vertices connected by edges and its length is given by the number of edges that connect the two extreme vertices;
- *Geodesic distance* consists in the number of edges that connect two vertices by the shortest path and is, usually, the most efficient way of connection between two actors. (Hanneman & Riddle, 2005)
- *Eccentricity* is the distance measured by the shortest path between a vertex and the vertex further apart from it, in other words, it is the geodesic distance between a vertex and the vertex further apart.

a. Actor-level measures

When an actor is very involved in the relationships around him, whether directly or indirectly, he is said to be a central actor, this is, an actor with relatively high values of centrality.

The three main measures used to analyze the centrality of a node are as follows:

- The *Degree* of a vertex is given by the number of edges that surround it and allows measuring the vertex’s involvement in the network.
  Related to the *degree*, in a direct network, there are two concepts that can be distinguished:
    - *In-degree* which corresponds to the number of connections that are coming into a specific vertex. If an actor receives many connections but starts just a few, i.e., if he has a high value of *in-degree* he may be considered a popular actor.

- *Out-degree* which corresponds to the number of connections that are going out of a specific vertex. If an actor has a high out-degree is considered an influential actor. (Hanneman & Riddle, 2005)

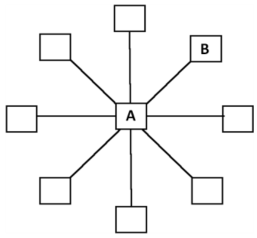

Figure 9 – Example of a Star Network

In **Figure 9**, the central actor – node A – has a degree centrality equal to 8 and its peripheral actors, such as node B, have a degree centrality equal to 1. From here we can see that node A is in a better position than all other actors since, for instance, if actor B cuts relationships with actor A, he would be isolated, while actor A would have 7 other possible contacts.

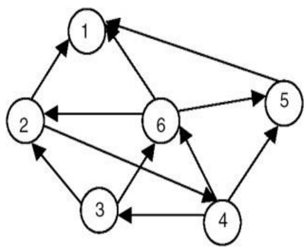

Figure 10 – Example of a direct graph

In **Figure 10**, regarding the in and out degrees, we can see that *node 1* only receives links and has an in-degree equal to 3, out of all 5-possible links, and has an out-degree equal to 0, since no connections start from it. *Nodes 4* and *6* are the ones with the higher out-degree, both equal to 3. We can, then, assume that *Node 1* is the most popular and that *Nodes 4* and *6* are the most influential.

- The *Closeness centrality* index represents the degree to which an actor is close to all others in a network. An actor that is close to many others in the network can in an easy and faster way communicate with them or access information more efficiently, since he will not need to resort to many intermediaries.

Closeness is given by:

$$Cc(n_i) = \frac{n-1}{Sum\ of\ the\ geodesic\ distance\ (n_i, n_j)}$$

Where:

- $Cc(n_i)$ represents the standardized closeness centrality of node i;
- n represents the number of nodes in the network;
- The *sum of the geodesic distance* (n_i,n_j) represents the geodesic distance between node i and all the other nodes.

Referring back to **Figure 9**, we can now see that node A has a closeness of 1 (since its geodesic distance to all the other actors is equal to 1, the sum of the geodesic distance is equal to 8, and (n-1) is also equal 8). On the other hand, if we are analyzing the closeness of node B, we can see that it is equal to 0,533 (its geodesic distance to all nodes except node A, which equals 1, is equal to 2, what results in a total of geodesic distance of 14 and (n-1) equals 8, so we have $\frac{8}{15} = 0{,}5333$).

- "Interactions between two nonadjacent actors might depend on the other actors in the set of actors, especially the actors who lie on the paths between the two." (Wasserman & Faust, 1994). *Betweenness* is an indicator that lets us know to what extent an actor plays the "broker" role in a network, this is, "examines the extent to which an actor is between all other actors within the network". (Everett & Borgatti, 2005)

  This indicator is given by:

$$C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}}$$

  Where:

  - $C_B(n_i)$ represents the normalized betweenness score of node i;
  - The numerator corresponds to the quantity of connections between *node j* and *node k* where *node i* acts as a broker
  - The denominator corresponds to all the possible intermediary connections.

Once again, based on **Figure 9**, we can say that node A is among all possible links so its betweenness equals 1. In contrast, node B is not part of any intermediation which means its betweenness is equal to 0.

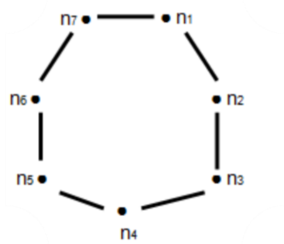

Figure 11 – Example of Circle Network

On the other hand, if we look to **Figure 11**, a circle shaped network, where there is no central node, we can see that each node has a betweenness of 0,2. There are 15 possible links where node *n1* acts as an intermediary but only in 3 of them he is the actual broker (through the shortest path), which are from *n6* to *n2*, *n7* to *n2* and *n7* to *n3*.

Besides the measures presented above, there is another indicator of the importance of a vertex, which is the *Eigenvector centrality*. This indicator is based on the idea that the power and status of an actor is dependent on the power and status of its neighbors. This measure, which can be seen as an extension of the measure of *degree*, takes into consideration not only the quantity of connections but also the quality of these connections. The term "eigenvalue" corresponds to the location of each actor considering its dimension and "the collection of such values is called the "eigenvector"." (Hanneman & Riddle, 2005) While the *in-degree* metric only considers the number of links that point towards a certain node, the *eigenvalue centrality* metric assigns relative scores to the nodes, according to the importance of its connections. With this metric, "the centrality of a given node *i* is proportional to the sum of centralities of its neighbors". (Oliveira & Gama, 2012)

As seen before, one characteristic of Social Networks is related to Transitivity. This concept can be measured through the *clustering coefficient*, which measures the degree to which vertices tend to cluster. If we are analyzing nodes, we use the *local clustering coefficient*, also known as *node clustering coefficient*; if analyzing the network, we use the *global clustering coefficient*. The *local clustering coefficient* of a vertex tells us how likely it is that two neighbors establish a connection and, in an indirect network, it is given by:

$$C_i = \frac{\#\, of\ links\ between\, neighboring\ vertices}{\#\, possible\ links\ between\, neighboring\ vertices}$$

In **Figure 12**, we can see three different networks. The colored node represents the actor whose local clustering coefficient we are calculating. In *Network A*, we have a network in which neighbors have no relation between them, so, in this situation, the clustering coefficient equals 0; in *Network B* there is one link between two of the neighbors which makes the coefficient to be equal to 0,33, since there is 1 link between the 3 possible; in *Network C*, all of the neighbors are connected so the local clustering coefficient equals 1.

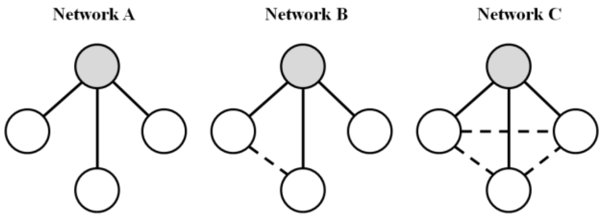

Figure 12 – Example of three different connections between neighbors

b. Network-level measures

In the previous point, we referred to the measures used when analyzing the way actors are positioned in the network, essentially at the level of their connections. Now, our focus will be on the structure of the network.

The *Density* of an indirect network is given by the proportion of ties that exist out of all possible ties and varies between 0 and 1. If it is a direct network, this index is given by the proportion of pairs that exist out of all possible pairs. When the network has a density near to 1, it is considered a dense network.

If we pay attention to **Figure 13**, we can notice that, in the social network B, there are more actors with just one connection – there are 6 – while in social network A there are only 2. It is then possible to deduce that Social Network A has a more active social life than Social Network B, in other words, A is denser than B.

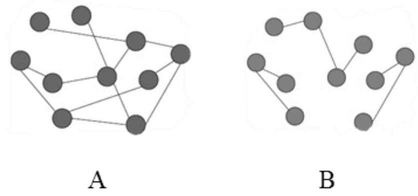

Figure 13 – Example of networks with different densities

The problem with this index has to do with the fact that it is sensitive to the number of nodes in the network; reason why it should not be used in comparisons across networks. (Ghali, Panda, Hassanien, Abraham, & Snasel, 2012)

If analyzing undirected networks, knowing the density of the network gives us all the necessary information about the network since the connections either exist or not. However, when analyzing directed networks, the situation changes. In this case, there are four possible scenarios: nodes 1 and 2 are interconnected, node 1 is connected to node 2, node 2 is connected to node 1 or they are not connected at all. When the actors are interconnected, it is said that they have a reciprocal connection. In this regard, there is a metric called *Reciprocity* concerning the probability of two vertices sharing the same type of connection.

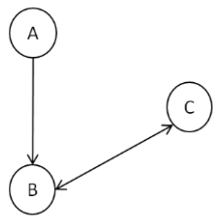

Figure 14 – Example of triad with different type of connections

There are two ways to calculate this indicator:

$$Reciprocity = \sum \frac{Reciprocal\ ties}{Possible\ pairs} \qquad \text{or} \qquad Reciprocity = \sum \frac{Reciprocal\ ties}{Pairs\ formed}$$

The second way is mostly used in large networks, since there are many actors that do not have a direct tie to many other actors. From **Figure 14**, with the first formula, the reciprocity of the network would be equal to 0,33; with the second formula, it would be equal to 0,5, since there is no connection between nodes A and C. (Hanneman & Riddle, 2005)

While in the actor-level measures there is an indicator that gives us a sense of the position that the actor occupies in his network – Centrality – at the network-level measures, there is a metric called *Degree of Centralization* which tells us at what level a network is dominated by one or more actors.

The *Degree*, in terms of network analysis, is a "measure of network activity or cohesion and is calculated using sociocentric data from the total number of ties to and from and individual" (Abbott, Bettger, Hampton, & Kohler, 2012). This measure allows us to know if the network has a centralized structure.

While Density tells us the level of connection in the network, the Degree enables us to know if the network is organized around some specific vertex/vertices.

If the value of the degree of centralization is equal to 0, it means that every node is connected to all other nodes; and if it equals 1 it means that the network develops around a specific actor – star shaped network. (Ghali, Panda, Hassanien, Abraham, & Snasel, 2012)

As previously mentioned, when we want to analyze the clustering coefficient of a network, we use the *global clustering coefficient*. This measure is based on triplets of nodes, which is a set of three nodes that are connected by two – open triplet – or three – closed triplet or triangle – undirected ties, and is given by:

$$C = \frac{number\ of\ closed\ triplets}{number\ of\ possible\ connected\ triplets}$$

In **Figure 15**, if we look at *node A* we see that it is connected to two nodes that are also connected to each other, so it has 1 closed triplet formed in its network in 1 possible triplet, which results in a global clustering coefficient equal to 1; *Node B*, in turn, has 3 connections, that is, 3 possible connected triplets (A-B-C, C-B-D and A-B-D) but only two closed triplets (A-B-C and C-B-D), which results in a global clustering coefficient of 0.67.

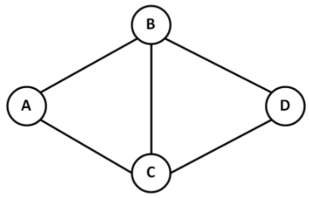

Figure 15 – Example of network to analyze global clustering coefficient

2.1.3.2 Ego Networks

In the previous topic, were presented some of the main metrics used to analyze Social Networks. However, not all of them are used when analyzing Ego Networks.

One of the measures presented before that can be applied to Ego Networks is the *centrality degree*, which tells us to what extent the Ego occupies a central position in its network.

The most relevant metric that allows us to know about the position of the Ego in the network is the *betweenness* which, as already mentioned, if equal to 1 means that the Ego plays the role of the broker in the whole network.

An important thing to know about Ego Networks is related to the cohesion of the network, this is, "the degree to which alters in the Ego Network are connected to each other" (Wielens, 2014).

The *local clustering coefficient*, also known as *Ego Network density*, is one of the measures used to study the cohesion of the network, this is, it allows us to know how connected are all the alters around the ego. If this value is near to 1, it is a very cohesive – dense – network.

Another measure that lets us know more about the connections between alters is the *global clustering coefficient*, which measures the transitivity of each actor. In this indicator, the Ego is excluded, since it is connected to every actor of the network and the point is to study the connection among alters. Another reason to exclude the Ego from the calculations of this metric is that it "simplifies calculations dealing with isolates, who would otherwise extremely skew any transitivity score". (Brooks, Hogan, Ellison, Lampe, & Vitak, 2014)

Knowing the *Reciprocity* allows us to know the extent to which the Ego's connection with the alters is mutual.

If, on the one hand, the existence of structural holes in an Ego Network might be considered negative since it means that there are some missing connections between the alters, and therefore the network isn't very cohesive; on the other hand, it might be benefic to the Ego since it means that the Ego has more autonomy and control over its alters.

2.1.4 Community Detection in Ego Networks

In our daily relationships, it is natural that we recognize certain people as belonging to a specific social group within our network, either because they are from our family, they studied in our university, are our neighbors or even because they work with us.

In online social networks, such as Facebook, it is possible to form groups or lists – social circles – with friends that share some characteristics, which, as said before, might be the city they leave in, the university they studied at, etc. These groups vary from person to person and an individual may also belong to more than one of the groups. (McAuley & Leskovec, 2013)

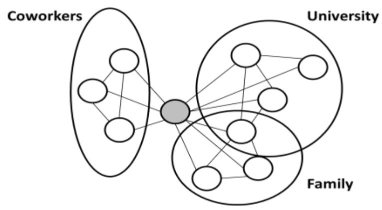

Figure 16 – Example of Social Circles

In **Figure 16**, assuming the colored node as being the Ego, we can see that there is an alter that belongs to two distinct groups – University and Family. When this happens, we say that it is a multiplex relationship. When an actor belongs to only one social group, it is said to be a uniplex relationship. Within each group may even exist a sub-group with higher density, for example, in the University circle there may exist a denser circle with the classmates.

The formation of these social circles makes us recall the concept of homophily, i.e., there is a tendency to group in the same social circle people who share certain characteristics since they have higher chances of establishing a connection. However, these *Facebook* social circles may not correspond to the communities detected in the network, when analyzing it with some community detection algorithms since the Ego can agglomerate in the same group people that, even though they play the same sport, for example, they do not know each other (Zafarani, Abbasi, & Liu, 2014); and, in a community, it is assumed that the majority of the actors are interconnected.

In an online social network and from a business perspective, it might be interesting to study the communities of a network because we will be, then, able to identify some groups of customers that visit our page and, consequently, target campaigns that best meet their needs and preferences or, for example, it also allows to see what kind of characteristics of the visitors are less explored by the products or services of the company. Additionally, it also lets us know which actors are in the center or in the periphery of each cluster since "vertices with a central position in their clusters, i.e., sharing a large number of edges with the other group partners, may have an important function of control and stability within the group" (Fortunato, 2010)

There are many algorithms whose focus is the detection of communities and that allow us to conclude on the characterization of its elements as well as its constitution, which would not be possible without this kind of analysis.

One of the issues related to the detection of communities is called graph partitioning and has to do with the division of the network into groups with almost the same size while also minimizing the number of edges that connect vertices of different groups (Newman, Detecting community structure in networks, 2004). Graph partitioning consists in dividing the network in two parts and then divide it until obtaining the ideal number of groups. This algorithm implies that we define a starting partition as well as its size.

Since this solution is quite iterative, there are other methods that allow to obtain this community division. One of them is the Kernighan-Lin algorithm, which is a "greedy optimization method that assigns a benefit function Q to divisions of the network and then attempts to optimize that benefit over possible divisions" (Newman, Detecting community structure in networks, 2004) being the benefic function the difference between the number of edges within two specific groups and the number of edges that connect those groups. Despite the efforts to improve it, the major negative point of this algorithm is its running time and the fact that it does not let us know when to stop the process of dividing the network.

One method that does not need us to specify the size of the groups and requires less running time, is the Girvan-Newman algorithm. This one is a hierarchical method that detects communities by iteratively removing edges from the network which were chosen according to its betweenness score. This score is re-calculated every time an edge is removed. The major negative points of this algorithm are that it is also time consuming and it does not tell when to stop removing edges. To surpass this disadvantage, Girvan and Newman suggested using the modularity index. (Newman, Detecting community structure in networks, 2004)

The most used community detection algorithm is then called *Modularity* which "is, up to a multiplicative constant, the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random" (Newman, Modularity and community structure in networks, 2006). Its values vary in a range between -1 and 1 and it can be considered a significant community structure when its values are higher than 0,3. If the value of modularity is high, it means that the density of the connections within each detected community is also high and the connections between nodes of different communities are sparse.

2.2 Trust

Although "trust" is a very present concept in our daily lives, it is not exactly easy to define and not everyone has the same definition since it can be seen from a philosophical, sociological or even economic perspective.

A trust relationship, regardless of its perspective, is always the interaction between, at least, two agents: the one who trusts – trustor – and the one who is trusted – trustee.

Below are some definitions presented by several authors about trust:

- "To rely upon a person to fulfil a commitment" (Hawley, 2014)

- “To believe that something is true although you have no proof” (Cambridge University Press, 2017)
- “Relation in which at least one party places a valued enterprise at risk to the malfeasance, mistakes, or failure of another party” (Tilly, 2010)
- “Firm belief in the reliability, truth, or ability of someone or something” (Oxford University Press, 2017)
- “Relationship in which a trustor decides to depend on the trustee’s foreseeable behavior in order to fulfil his expectations” (Turilli, Vaccaro, & Taddeo, 2010)

Although its definition is not precise – “trust means different things to different people, to different roles, and in different scenarios” (Sisson, 2017) – there are some characteristics common to all definitions. Taking this into consideration, in his dissertation, Bo Fu (2007) compiled the main characteristics of Trust:

- Asymmetry: “when a person is ill and goes to the doctor the information is asymmetric because it supposes the patient’s acceptance and belief on medical advices, in view of the seriousness of the situation, so the patient needs to trust” (Kuz & Giandini, 2012)
- Transitivity: If I trust my friend 100% and he tells me he trusts his brother 100%, then it is quite probable that I trust my friend’s brother as well;
- Subjective: As mentioned before, trust does not have the same definition to everyone. When talking about politics or even sports there are always different people who consider certain people more trustworthy than others;
- Context-dependent: We may not always trust someone in every situation; with some people, we only trust commonplace issues.

As mentioned, a social network consists of a graph in which are represented some entities – nodes – and their connection – edges. When the relation represented in the graph is related to trust, then this network is called “Trust network”.

Similarly to every social network, trust networks can be “offline” social networks – networks that depicts day-to-day connections without resorting to the internet – or “online” social networks – networks which are formed over the internet; however, in a trust network the edges are directed since A can trust B but B does not necessarily have to trust A. (Golbeck, Parsia, & Hendler, 203)

So, although "trust" can be analyzed from several points of view and according to different fields of study, our focus will be on trust networks in online social networks, which is the subject of the next sub-topic.

2.2.1 Trust in Online Social Networks

From an interpersonal point of view, trusting someone we know is not that difficult; trusting our parents to take care of us or our friends when we tell them something private is not that complicated, because we know them well and we know to what extent we can trust them.

Nevertheless, trusting someone online that we do not know is not that simple at all since we know nothing about them and, therefore, we are not able to understand their level of trustworthiness in a particular situation. That is why "we only use information found on the web when we trust the source". (Victor, Cornelis, & De Cock, 2006)

In his work, Volakis (2011) mentions two conditions underlying the existence of trust, which are:

- "The participants should possess as far as possible a common background considering cultural and institutional aspects and
- One participant should be sure about the other's identity."

However, this kind of information is not always shared, most often for privacy reasons, so people need to rely on other characteristics, such as reputation.

An example of an online social network which relies on trust is a recommendation website. People often turn to these websites for information they do not have, either to know people's opinion about a restaurant or even the best store to buy a certain product. However, not everyone is trustworthy so people need to determine if they should trust them or not.

Most recommendation websites allow us to assign rates to the products under review and some of them even allow us to assign rates to the people who reviewed the products. In order to determine whether or not to trust a particular person, we may try to answer a few questions:

- Is this person’s opinion well-ranked?
- Have the reviews of this person been recommended to me?

The answer to the first question is often related to a person’s reputation. The reputation of a person indicates “how a particular agent is expected to behave” (Josang, Hayward, & Pope, 2006). The higher the score, more likely it is that its reviews will be accepted by the majority, this is, the higher the reputation of an individual, more likely he is to have people trusting him.

The second question is about transitivity. If A trusts B that trusts C, and B tells A that he trusts C, then A might consider trusting C as well. By telling A that he trusts C, B is making a recommendation. (Josang, Hayward, & Pope, 2006) This recommendation situation, however, is easier to come by personally. Online it is more likely to happen based on a recommendation system, this is, if A trusts B that trusts C, it is possible that a suggestion appears to A to also trust C or even that A begins to see more often reviews from C.

If we tried to build a trust network based on the trust relationships of these recommendation websites, the most likely would be that people with higher reputation would also have more people trusting them, this is, would have more edges directed to them; and people who would appear in the recommendations would be those who were closer to the node but not directly connected.

2.3 Other works

In this topic, we will mention some works in which the focus are social network analysis, particularly Ego Networks, as well as works related to trust networks, all of them related to online social networks.

According to Robin Dunbar (Dunbar, How many friends does one person need?, 2010), a person can establish a social relationship, where each person knows each other reasonably well, with no more than approximately 150 people – this being designated as Dunbar’s number. Of these 150 people, we cannot relate to all of them in the same way.

Dunbar divides them by layers. In the nearest to the Ego, in the inner-core, there are about 5 people; in the next one we can have around 15 people (including the initial 5 people); in the third layer, there may already be up to 50 people and after that we can have the 150. Dunbar still adds that there may be at least two more layers after that, "one at 500 which you might think of as acquaintances, so again this is including everybody within the 150 as well; finally, one at 1500 who are basically the number of faces you can put names to". (Dunbar, Robin Dunbar on Dunbar Numbers, 2013) For a better understanding of this layered division, a return to **Figure 7** is suggested.

In their work, Arnaboldi *et. al* (2012) intend to realize if these Dunbar's social circles exist in the same way in online social networks. To do so, they analyze a large dataset from Facebook with more than 23 million social interactions and compare it to Dunbar's theory. They found that in online Ego Networks the number of social relationships of each type – each layer – and the average size of the network was very close to Dunbar's number. They state that even though online social networks allow us to reach more people, people tend to establish relationships with the same criteria as they do in offline social networks.

Similarly, by choosing what we want to share with other people, such as opinions, beliefs or even ideas, we indirectly determine how others see us and how they relate to us; that's why people tend to "trust more those who they know than those who they don't". (Kuz & Giandini, 2012)

Many times, in online social networks such as Facebook, for instance, people usually relate first to the people they already know in everyday life but they also establish relationships because they are suggested to them, either because they have friends in common – transivity principle – or because they have very similar preferences.

Authors such as Guha *et. al* (2004), Rad *et al.* (2012), Zolfaghar *et al.* (2010) developed works whose focus are trust networks in online social networks. On their work, Adali *et al.* (2010) focused on quantifying "behavioral trust", which is the trust "based on observed communication behavior in social networks"; by analyzing the frequency of Twitter conversations, through metrics they developed over timing and sequence of conversations, they were able to see that these two metrics were highly correlated with each other and that they also correlated to a behavior indicative of trust. Zolfaghar *et al.* (2010) define this “behavioral trust” as being a “knowledge-based trust” which is the kind of trust that comes from the level of satisfaction that an individual derives from the relationship with another individual; if individual A has a high level of satisfaction related to individual B, then A knows that the probability that individual B will disappoint him is small; the more interactions A and B have, the more knowledge the first has over the second. Zolfaghar *et al.* (2010) also define “reputation-based trust”, which can be based on the popularity of an individual on a certain topic, such as recommending places where to go for dinner.

3 Case study

Considering everything that has been introduced about Trust and Social Networks, we can now analyze the concrete case of Epinions trust network, using tools such as *Gephi* and *Excel*.

We will begin with a brief contextualization of Epinions, followed by a global analysis of the social network, where, besides analyzing the main metrics of the network, we will identify two Ego Networks to analyze based on a certain criterion. Next, we will analyze those Ego Networks and the impact of the presence / absence of each Ego in its own Network.

The purpose of this analysis is to measure the impact that the Ego has on the structure of its trust network, if it has significant influence on the formation of connections in the network as well as on their maintenance.

3.1 Epinions

As previously mentioned, this work will be based on a trust network extracted from *Epinions.com,* therefore we consider it important to make a brief contextualization regarding the functioning of this website.

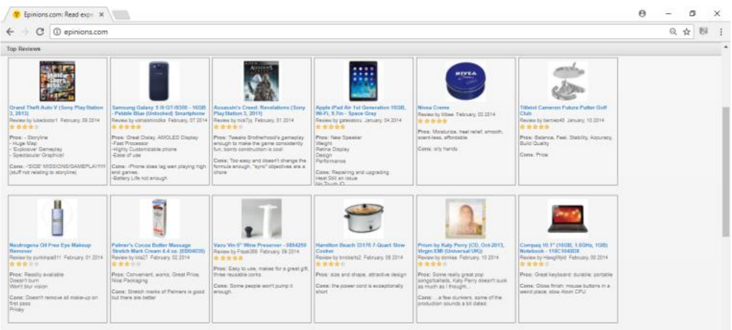

Figure 17 - Printscreen of the website *Epinions.com*

Epinions was a website where people could review several products that could belong from the electronic to the interior decoration sectors, , as we can see by **Figure 17** which is a printscreen of the home page of *Epinions.com,* with the particularity that users could create a list with the people on whose reviews they trusted the most – thus forming their trust network.

In addition to the positive effect that the creation of this trust network had for the user, who was able to see more easily the reviews of the people he considered most reliable, this list would also influence the people that had this user on their trust network, thus creating a potential influence on the network of the other users. (Epinions.com, 2017)

Our dataset is, therefore, an extraction of these trust networks collected by Paolo Massa between January 2001 and August 2003. It consists of a direct network with 95.318 nodes and 841.372 edges where each node represents a user and each edge represents a trust/distrust connection and it also gives us information about the date the connection was established. Since it is a direct network, there may be two edges connecting the same two nodes. (Massa & Avesani, 2007)

For the purpose of our analysis, we will only consider trust connections, disregarding, thus, the distrust connections. We will also begin by analysing the network of 2003 – because it is smaller and, therefore, its analysis takes less time and less storage space – and will only resort to the connections made in 2001 and 2002 years to analyse them in a dynamical approach.

In the following sections, we will analyse the global network of 2003 – its actor-level and network-level metric. Then, based on a certain criterion, we will select two different nodes to analyse their trust Ego Networks, and we will also analyse the communities' structure throughout the years and all the metrics.

Therefore, our initial trust network is a direct network with 8.719 nodes and 31.647 edges, which means 8.719 users and 31.647 trust links.

3.2 Social Network Analysis

Through *Gephi*, by importing our database and using Yifan Hu Proportional layout[1], we were able to get the view of our trust network, presented in **Figure 18**.

[1] We used this layout because it performed better than others due to our hardware limitations.

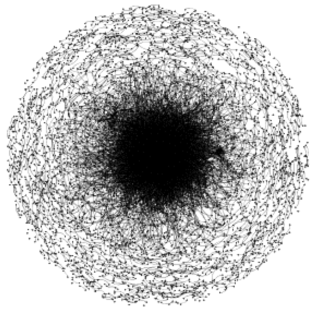

Figure 18 – Epinions 2003 trust network

Gephi also allows us to get some of the key social network analysis metrics, which are presented in the table below.

Metric	Value
Average degree	3,63
Density	$8{,}706 \times 10^{-6}$
Network diameter	11

Table 1 – Epinions network metrics

Based on the definitions presented in the sub-chapter "2.1.2 Metrics" we can understand the values in **Table 1**, which will now be interpreted.

An average degree equal to 3,63 means that, in average, each node is connected to 3,63 other nodes. This value in a network with 8.719 nodes, leads us to believe that the network is not very interconnected, this is, that the network is a dispersed network, fact proven by the value of density very close to zero.

Considering that the distance between two nodes is the number of connections through the shortest path that exists between them and that the network diameter is the maximum distance between any two nodes in the network, knowing that the network diameter of our graph equals to 11 means that the largest component of the graph has, among its more distant nodes, a path of 11 edges.

Gephi also allows us to export a document which contains the statistics of each node of the network, so we can know, for instance, which are the nodes with higher and lower in and out-degree. Thus, to proceed our analysis of Ego Networks, we will select a node according to its statistics.

Based on the in-degree metric which, as previously mentioned, corresponds to the number of edges that are coming to a specific node and gives us a sense of the popularity of the actor, we will choose the ego with the highest value, which is node 0103. We will also analyze a node based on the out-degree metric, which indicates the influence of a node, which is node 0003.

3.3 Ego Network Analysis

3.3.1 **0103 Ego Network**

3.3.1.1 Network-level analysis

Filtering our social network so that we can only visualize the Ego Network of node 0103 we get the graphs presented in **Figure 19**, according to the depth we choose. To achieve these graphs, we used Yifan Hu's layout, once it was the layout that gave us a better representation of each network.

In *Gephi* we can define an Ego Network with depth 1, that only shows the connections made directly to and from the ego as well as the connections between the alters – we can see the friends of the ego and the relations of friendship among the Ego's friends; with depth 2, we can see the friends of the ego and the direct friends of the friends – the acquaintances of the ego; when the depth equals 3 we see not only the friends of the ego but also their friends' acquaintances. So, it is expected that the graph increases as its depth increases as well – not only in the number of nodes but also the edges. On the other hand, it is expected that the density of the graph as well as the average degree decreases; since with the increase of the nodes the connections that arise do not grow proportionally, it is expected that there will be more nodes but less connections between them.

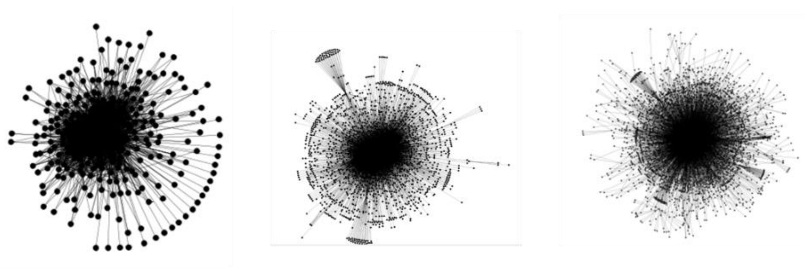

Figure 19 – Node 0103 trust network

In **Table 2** we present some of the main network-level metrics of each of the graphs. Node 0103's Ego Network will be designated as Graph A followed by ".1", ".2" or ".3" according to the depth of the graph.

Metrics	Graph A.1	Graph A.2	Graph A.3
Nodes	315	2373	4695
Edges	3633	25240	28844
Average degree	11,533	10,636	6,144
Density	0,037	0,004	0,001
Network diameter	5	8	10
Modularity	0,098	0,177	0,236
Connected components	1	1	1
Avg. Path length	2,227	3,385	3,899

Table 2 – Node 0103's Ego Network metrics

As we expected, the number of nodes and edges has a significant growth which also results in an increasing of the diameter of the network, since to go from a node to its most distant node, it will now be necessary to travel a longer path.

3.3.1.2 Node-level analysis

With regard to the analysis at the node-level, the two main metrics we are going to focus on are *closeness centrality* and *betweenness centrality*. Since we are dealing with a direct network, we will also take into consideration the information about the strong connected components of the network.

As we have seen, the *closeness centrality* gives us a sense of the closeness of a certain node to all the other nodes in the network and, the lower this value, the further away the node is from the generality of the network. Relative to node 0103, being the Ego, it is expected that he has the highest value of *closeness* as well as of *betweenness centrality*, since this metric lets us know how often a specific node acts as an intermediary, through the shortest path, between the other nodes.

ID	In-degree	Out-degree	Closeness	Betweenness
0103	265	199	0,773	0,545
0003	2	103	0,611	0,007
1493	35	71	0,555	0,009
0676	24	67	0,548	0,009
0017	56	59	0,545	0,017
1028	29	57	0,537	0,006
0298	23	50	0,533	0,008
1049	28	48	0,532	0,007
0714	33	47	0,531	0,012
0028	28	53	0,528	0,009

Table 3 - Graph A.1 Top closeness

In Graph A.1 as expected, the node with higher values of *closeness degree* and *betweenness degree* is the Ego, node 0103, with 0,773 and 0,545, respectively, since this graph only shows us the connections made directly from and to the Ego. This means that the Ego is in a very central position, as it was expected, with easy access to the other nodes of the network.

As we can see in **Table 3**, the second node of this network with the highest value of *closeness* is node 0003, whose value equals 0,611, followed by node 1493, which means that, after the Ego, these are the nodes with the most central position of the network.

ID	In-degree	Out-degree	Closeness	Betweenness
0103	265	199	0,773	0,545
0196	35	38	0,526	0,019
0017	56	59	0,545	0,017
0201	59	28	0,493	0,014
0032	38	33	0,507	0,013
0714	33	47	0,531	0,012
0147	45	36	0,509	0,011
0711	22	24	0,492	0,011
1061	28	35	0,509	0,010
0631	11	14	0,479	0,010

Table 4 - Graph A.1 Top betweenness

Table 4 shows us that the node with the highest value of *betweenness* after the Ego, is node 0196, which presents a much smaller value of 0,019, followed by node 0017; this means that, compared to the Ego, these nodes act much less as a bridge. It should be noticed that, since this is a direct network, not every node is accessible. In fact, in Graph A.1 there are 45 nodes with in-degree equal to 0; the same way, there are 20 nodes with out-degree equal to 0.

In **Figure 20** we structured Graph A.1 using the layout *Fruchterman Reingold* since it gives us a more pleasant and less condensed view of the graph, and we highlighted the Ego (colored pink), the nodes with the highest values of *betweenness* (colored green) and the ones with the highest values of *closeness* (colored blue).

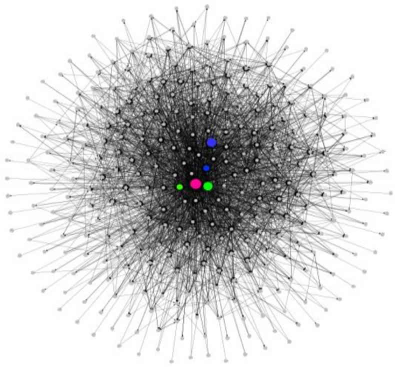

Figure 20 - Graph A.1 Top nodes

In Graph A.2, since there is a higher number of nodes and connections, it is expected that the values of *betweenness* are smaller, because there are more nodes to whom each node can connect to and, thus, doesn't need to depend on a specific node to maintain itself in the network. The same way, to what concerns to the *closeness* metric, it is expected that, once there are more possible paths, there might be more nodes with high values of this metric.

The node with the highest *betweenness* is the Ego followed by node 0192, with values of 0,071 and 0,036, respectively, as can be seen in **Table A. 1**.

As can be seen in **Table A. 2**, there are twelve nodes with *closeness* equal to 1, however, as we've mentioned before, being this graph a representation of a direct network, it is important to complement our analysis with information about the strongly connected components since, this way, we will know if a closeness equal to 1 really means that the node is connected directly to all the other nodes or if he is part of a small strongly connected component. In order to determine the largest connected component of our network, we used *Gephi* to identify Graph A.2's components. We then filtered through the largest component of the network and noticed that the node with highest value of *closeness* is node 0003, with a value of 0,510. The Ego has a *closeness* of 0,433, which means that, increasing the depth of his ego network, the Ego decreases its central position.

In Graph A.3 we expect that there are more strongly connected components and the values of *betweenness* to be even smaller. In fact, there are now 3142 strongly connected components, while in Graph A.2 there were only 1104 and in Graph A.1 65. In this graph, as we can see by **Table A. 3**, the node of the biggest strongly connected component with higher value of *closeness* is node 0003, with *closeness* equal to 0,457, followed by node 0101 who has *closeness* equal to 0,439; and the node with higher value of *betweenness* is node 0192, with 0,030, followed by the Ego who has a *betweenness* of 0,028 and a *closeness* of 0,379. This means that, once again, the Ego lost its position as the most central actor of the network and, additionally, is no longer the one that most acts as an intermediary.

3.3.1.3 Dynamic Analysis

In this sub-topic, we will analyze the development of the network regarding the metrics of the nodes that surround the Ego and the Ego itself, throughout the three years.

This analysis will only be done based on a depth equal to 1 and we will select three different moments in time and then calculate the metrics of the nodes.

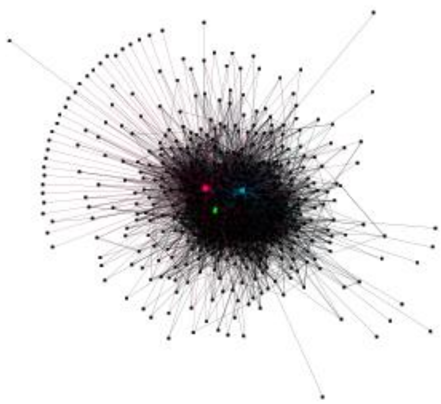

Figure 21 - Moment 1

The first moment we selected was the time interval between Jan 1^{st}, 2001 and Feb 28^{th}, 2001. At this moment, the network has 556 nodes and 9.324 edges. The node with the highest value of in-degree, equal to 392, is the Ego (pink node in **Figure 21**), followed by node 0297 (blue node), with in-degree equal to 208, and node 6037 (green node). In **Table 5** we present some of the metrics of the nodes with higher in-degree.

ID	In-degree	Out-degree	Closeness	Betweenness
0103	392	174	0,507	0,322
0297	208	113	0,444	0,065
6037	114	50	0,389	0,011
4378	104	26	0,371	0,005
2589	104	16	0,373	0,002
3105	103	15	0,342	0,001

Table 5 – Moment 1 - Top nodes metrics - higher in-degree

The second interval of time was the period extended to Feb 28th, 2002. At this moment, the network had 621 nodes and 13.700 edges. The Ego was still the node with highest in-degree.

ID	In-degree	Out-degree	Closeness	Betweenness
0103	526	227	0,549	0,320
0297	273	152	0,479	0,054
3163	143	63	0,427	0,007
6037	140	53	0,398	0,007
2936	121	58	0,427	0,006
2837	121	58	0,427	0,006

Table 6 - Moment 2 - Top nodes metrics - higher in-degree

As we can see in **Table 6**, between Feb 28th, 2001 and Feb 28th, 2002 there was a node that started stablishing new connections – node 3163. In the first moment of analysis it had an in-degree equal to 82. In **Figure 22** we can see the difference between this node in moment 1 and moment 2, the network became denser.

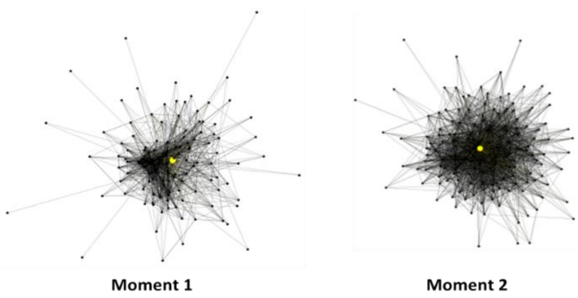

Figure 22 - Node 3163 Moment 1 *vs.* Moment 2

The third moment is until Aug 31st, 2003 and, at this moment, the network has 556 nodes and 12.687 edges. At this point, there were no changes in the nodes with the highest values of in-degree. The Ego is still the node with the higher value of in-degree, as well as of *closeness* and *betweenness degree,* as we can see in **Table 7**.

ID	In-degree	Out-degree	Closeness	Betweenness
0103	560	232	0,553	0,328
0297	287	161	0,494	0,054
3163	155	68	0,432	0,007
6037	141	53	0,402	0,007
2936	137	63	0,432	0,006
2837	133	60	0,420	0,010

Table 7 - Moment 3 - Top nodes metrics - higher in-degree

3.3.2 **0103 Ego Network (without the Ego)**

3.3.2.1 Network-level analysis

Using *Gephi*'s capabilities, we are also able to obtain node 0103's Ego Network excluding the node's own presence and, consequently, its connections, which results in the graphs represented in **Figure 23**.

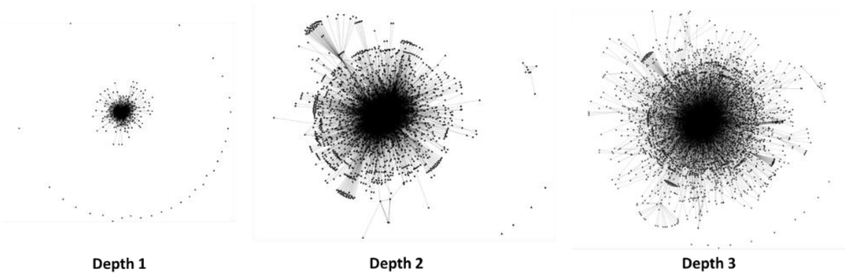

Figure 23 - Node 0103 trust network without node 0103

In **Table 8** we present some of the main network-level metrics of each of the graphs. Node 0103's Ego Network will be designated as graph A and Node 0103's Ego Network without the ego itself, will be designated as graph B.

Metrics	Graph B.1	Graph B.2	Graph B.3
Nodes	285	2372	4694
Edges	3169	24776	28380
Average degree	10,092	10,445	6,046
Density	0,032	0,004	0,001
Network diameter	7	8	10
Modularity	0,126	0,166	0,248
Connected components	30	13	13
Avg. Path length	2,821	3,435	3,925

Table 8 - Node 0103's Ego Network without the ego metrics

Unlike what happened when the Ego was present, without the Ego the network becomes not only sparser but with less connections among the nodes. While with the Ego there was only one connected component, without the Ego, when the depth of the graph is 1 (Graph B.1 of **Table 8**) there are 30 connected components, which means that the Ego was acting as a bridge to, at least, 30 nodes. As we increase the depth, the number of components decreases, since there will be more nodes and, consequently, more edges connecting these nodes. Nevertheless, even with depth 3 we can see that there are some nodes that are still isolated because they were depending of the Ego to connect them to the rest of the network.

3.3.2.2 Node-level analysis

In this sub-topic, we will not only refer to the nodes with the highest values of both metrics but also to the nodes we have highlighted in sub-topic 3.3.1.2.

As the Ego has been excluded, it will no longer be the node with higher values. In fact, its absence, as we have seen in the previous point, has an influence on the structure of the network and, therefore, it is expected that other nodes will stand out.

Regarding the *closeness* degree, it is not expected that these values increase, since no new connections will be created, in fact what might happen is that the values decrease since they will lose the connection to the Ego and the nodes to whom only the Ego linked them. On the other hand, the values of *betweenness* should have a slight increase since with the absence of the Ego there will be some nodes performing more of the intermediary function.

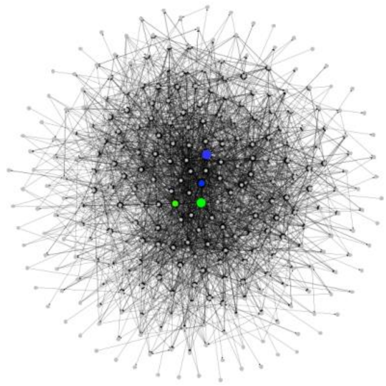

Figure 24 - Graph B.1 Top nodes

In Graph B.1, the nodes with higher values of *closeness* are node 0003 and node 1493, with 0,607 and 0,519, respectively, the same nodes as in Graph A.1, with slightly smaller values. The nodes with the highest values of *betweenness* are node 0017, with 0,068, and node 0147, which has as value of 0,042. These nodes, in Graph A.1, had *betweenness* of 0,017 and 0,011, as we can see in **Table 3** and **Table 4**, which means that, without the presence of the Ego and, therefore, the nodes that were only connected to him, nodes 0017 and 0147 act more as bridges. These variations are shown in **Table 9**.

ID	Closeness A1	Betweenness A1	Closeness B2	Betweenness B2	Var. Closeness	Var. Betweenness
0003	0,611	0,007	0,607	0,0222	-1%	217%
1493	0,555	0,009	0,519	0,0402	-6%	347%
0196	0,526	0,019	0,494	0,0404	-6%	113%

Table 9 - Nodes 0003, 1493 and 0196 Graph A.1 vs. Graph B.1

In Graph B.2, comparing to Graph A.2, there aren't much changes related to the *closeness* degree or the *betweenness* degree, as we can see in **Table 10**.

ID	Closeness A1	Betweenness A1	Closeness B2	Betweenness B2	Var. Closeness	Var. Betweenness
0003	0,510	0,034	0,509	0,036	-0,20%	5,88%
0101	0,478	0,022	0,48	0,025	0,42%	13,64%
0192	0,467	0,036	0,468	0,041	0,21%	13,89%
1493	0,450	0,022	0,448	0,024	-0,44%	9,09%

Table 10 - Nodes 0003, 0101, 0192 and 1493 Graph A.2 vs. Graph B.2

Isolating the nodes with the highest values of *closeness* and *betweenness* of Graph B.1 and comparing them with the values they now present, in Graph B.2, we realize that there is a node that stands out in terms of acting as a bridge, which is node 0003, as we can see in **Table 11**.

ID	Closeness_B1	Betweenness_B1	Closeness_B2	Betweenness_B2	Var. Closeness	Var. Betweenness
0003	0,607	0,022	0,509	0,036	-16%	64%
1493	0,519	0,04	0,448	0,024	-14%	-40%
0676	0,519	0,024	0,446	0,01	-14%	-58%
0017	0,518	0,068	0,439	0,028	-15%	-59%
0298	0,502	0,032	0,44	0,031	-12%	-3%
1028	0,499	0,016	0,425	0,005	-15%	-69%
1049	0,495	0,019	0,413	0,005	-17%	-74%
0196	0,494	0,04	0,439	0,016	-11%	-60%
0714	0,493	0,032	0,421	0,008	-15%	-75%
0028	0,486	0,03	0,418	0,017	-14%	-43%

Table 11 - Graph B.1 Top 10 closeness vs. values from Graph B.2

The other nodes decrease their values since the number of nodes and connections of the network increases greatly and other nodes assume more central positions, such as nodes 0101 and 0192 (**Table A. 4** and **Table A. 5**).

In Graph B.3, being the network with an already widened depth – depth 3 – it is expected that the presence or absence of the Ego does not have a significant impact on the metrics at the node level. And we can confirm this consulting **Table A. 6** and **Table A. 7**, where we can see the variation of each metric from Graph A.3 to Graph B.3.

3.3.2.3 Dynamic analysis

Just like we did in the sub-topic 3.3.1.3 we will now analyze the network in a dynamic perspective, this time excluding the Ego. The periods of analysis will be the same ones we previously used – Jan 1st, 2001 to Feb 28th, 2001; Jan 1st, 2001 to Feb 28th, 2002; Jan 1st, 2001 to Aug 31st, 2003.

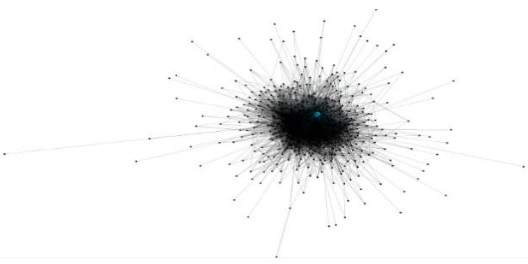

Figure 25 - Moment 1 - without the Ego

In the first period, the graph has 523 nodes and 8.758 edges and, as we can see in **Table 12**, the node with the higher value of in-degree is node 0297 (in blue), which was predictable since it was the node with the second highest value when we were considering the Ego in the analysis.

ID	In-degree	Out-degree	Closeness	Betweenness
0297	207	112	0,433	0,116
6037	113	50	0,388	0,024
2589	103	15	0,352	0,004
4378	103	25	0,352	0,012
3105	102	15	0,339	0,005
18416	99	67	0,390	0,017

Table 12 - Moment 1 w/o the Ego – Top nodes metrics – higher in-degree

In the second moment of analysis, the number of nodes increases to 577 and the number of connections is now 12947. **Table 13** shows us the values of in-degree, out-degree, closeness and betweenness of the top 6 nodes of the network in the second moment.

ID	In-degree	Out-degree	Closeness	Betweenness
0297	251	143	0,472	0,109
6037	133	53	0,405	0,020
3163	122	59	0,413	0,015
3105	114	21	0,365	0,005
2589	114	18	0,377	0,004
2936	113	57	0,418	0,012

Table 13 - Moment 2 w/o the Ego – Top nodes metrics – higher in-degree

In the third moment, there are a few new connections that are established. The graph has now 590 nodes and 14.119 edges and node 0297 is still the node with the highest value of in-degree, as shown in **Table 14**.

ID	In-degree	Out-degree	Closeness	Betweenness
0297	286	160	0,487	0,092
3163	154	67	0,419	0,013
6037	140	53	0,401	0,013
2936	136	62	0,419	0,009
2837	132	60	0,421	0,012
0002	129	81	0,443	0,014

Table 14 - Moment 3 w/o the Ego – Top nodes metrics – higher in-degree

3.3.3 **0003 Ego Network**

Just like we did to analyze node 0103's Ego Networks, we also filtered our social network in order to only get the Ego Network of node 00003, which is the node with the highest value of out-degree.

3.3.3.1 Network-level analysis

In **Figure 26** we can see the graphs of this Ego Network, according to the depth we choose.

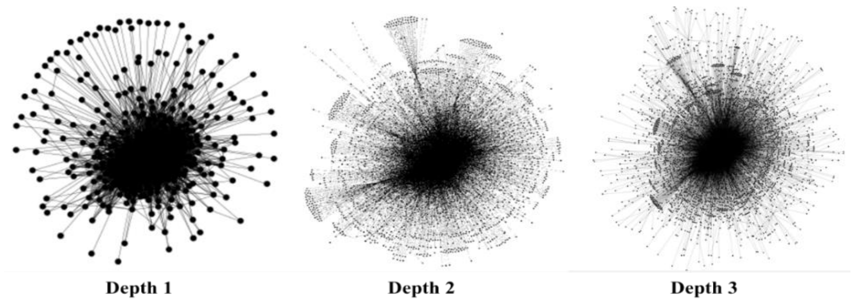

Figure 26 - Node 0003 trust network

In **Table 15** we present some of the main network-level metrics of each of the graphs. Node 0003's Ego Network will be designated as Graph C followed by ".1", ".2" or ".3" according to the depth of the graph.

Metrics	Graph C.1	Graph C.2	Graph C.3
Nodes	405	2913	4865
Edges	6584	26349	29039
Average degree	16,257	9,045	5,969
Density	0,04	0,003	0,001
Network diameter	6	8	10
Modularity	0,091	0,204	0,253
Connected components	1	1	1
Avg. Path length	2,744	3,482	3,941

Table 15 – Node 0003's Ego Network metrics

Once again, the number of nodes and edges increases significantly, as well as the diameter of the network and, consequently, the average path length.

On the other hand, the average degree has decreased due to the increase in the number of nodes that increased in a proportion higher than the number of edges.

3.3.3.2 Node-level analysis

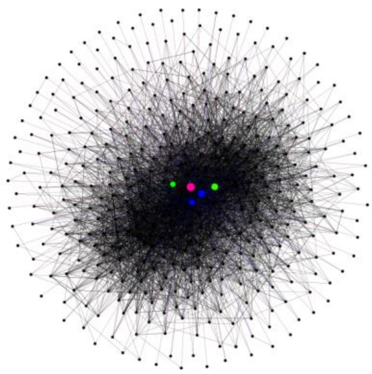

Figure 27 - Graph C.1 Top nodes

In Graph C.1, node 0003, being the Ego, is the node with the highest value of *betweenness*, equal to 1, and *closeness*, equal to 0,205. A *betweenness* equal to 1 means that node 0003 can reach all 404 nodes belonging to this network, which we can confirm by looking at **Table 16**. This means that the Ego is the most central node in the network.

ID	In-degree	Out-degree	Closeness	Betweenness
0003	15	404	1,000	0,205
0101	20	135	0,600	0,038
0092	19	81	0,556	0,019
0073	3	58	0,539	0,001
0152	18	53	0,535	0,015
0192	75	131	0,525	0,061
0231	36	38	0,525	0,047
0196	36	36	0,523	0,042
0148	12	31	0,520	0,011
0121	23	26	0,517	0,025

Table 16 - Graph C.1 Top closeness

On the other hand, the value of *betweenness* of the Ego is not that high, which means that even though he is the most central node of the network, he does not act as an intermediate for the connections of most nodes.

The nodes with the highest values of *betweenness* after the Ego, are node 0192 and node 0103, which present much smaller values of 0,061 and 0,056, respectively.

In **Figure 27** we highlighted the top nodes. The Ego in pink, node 0192 and 0196 in green and nodes 0101 and 0092 in blue.

In Graph C.2, with the increase of the number of nodes and edges it is expected that the *betweenness* of the Ego decreases significantly, as well as its position comparing with other nodes since he does not have that much direct links to him, as we can confirm by **Table 17**. It is expected that nodes with higher values of in-degree stand out, since this is the only way they can mediate a connection between two nodes.

ID	In-degree	Out-degree	Closeness	Betweenness
0092	161	316	0,465	0,041
0103	265	199	0,412	0,040
0101	25	308	0,485	0,034
0003	15	404	0,536	0,032
0017	140	172	0,436	0,023
0015	158	235	0,420	0,022
1493	109	243	0,441	0,022
0204	156	128	0,400	0,020
0150	38	145	0,383	0,019
0595	33	171	0,432	0,019

Table 17 - Graph C.2 Top betweenness

As we can see, compared to **Table A. 8**, the Ego is no longer the node with the higher value of *betweenness*.

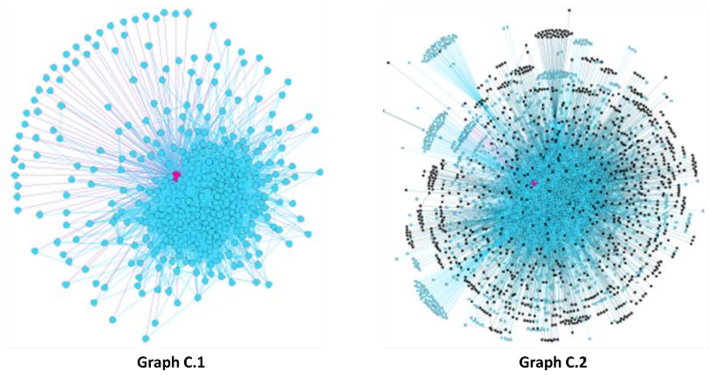

Figure 28 - Node 0003's neighborhood (Graph C.1 *vs.* Graph C.2)

In **Figure 28** we compare Graph C.1 with Graph C.2, regarding the Ego's neighborhood. The nodes painted in blue are the nodes that the Ego can reach and, as we can see, in Graph C.1 the Ego reaches every node of the network, as we have already mentioned, which explains the *closeness degree* to be equal to 1. On the other hand, in

Graph C.2 there are some nodes painted black, which means that the Ego does not have any contact with them. This lets us deduce that the *closeness degree* of the Ego will be significantly lower than it was. We can also confirm this by the values presented in **Table 18**.

ID	In-degree	Out-degree	Closeness	Betweenness
0003	15	404	0,536	0,032
0101	25	308	0,485	0,034
0092	161	316	0,465	0,041
0676	77	228	0,446	0,012
0092	24	131	0,446	0,005
1493	109	243	0,441	0,022
0017	140	172	0,436	0,023
0595	33	171	0,432	0,019
0152	40	129	0,429	0,007
0443	5	138	0,429	0,002

Table 18 - Graph C.2 Top closeness

In Graph C.3, even though the number of nodes has had an increase of 1101% comparing to the number of nodes of Graph C.1, the number of nodes the Ego connects to has remained the same. That means that its *closeness degree* and *betweenness degree* metrics will be smaller, as we can see in **Table A. 9** and **Table A. 10**.

3.3.3.3 Dynamic analysis

From Jan 1st, 2001 to Feb 28th, 2001, node 0003's Ego Network had 299 nodes and 4321 edges being the Ego the node with higher out-degree.

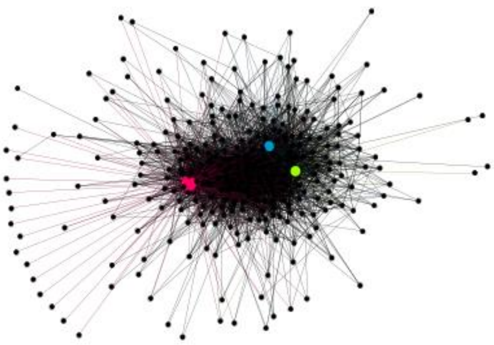

Figure 29 - Moment 1

In **Figure 29** we have highlighted the Ego, in pink, node 2998, in blue, and node 8440, in green, which are the nodes that present the highest values of out-degree - **Table**

19 –, that is, that are the nodes that usually have more influence on the connections established in the network.

ID	In-degree	Out-degree	Closeness	Betweenness
0003	107	108	0,578	0,231
2998	12	102	0,594	0,046
8440	5	91	0,572	0,009
8866	8	84	0,576	0,020
0076	49	80	0,529	0,052
1158	8	75	0,526	0,014

Table 19 - Moment 1 - Top nodes metrics – higher out-degree

In the second moment of analysis, until Feb 28th, 2002, our graph has now 345 nodes and 8.522 edges, which means that new connections have been established. In this case, where we are basing our analysis on the out-degree level, having an increase in these values means, as we will see in the sub-chapter 3.3.5, that between Feb 28th, 2001 and Feb 28th, 2002, there were more users of Epinions website that trusted the Ego, he only had 108 connections directed to him and has now 190, as we can see in **Table 20**.

ID	In-degree	Out-degree	Closeness	Betweenness
0003	190	190	0,689	0,256
2684	4	170	0,674	0,008
1158	27	170	0,667	0,019
2703	1	146	0,639	0,004
2998	21	136	0,611	0,023
0076	69	99	0,553	0,029

Table 20 - Moment 2 - Top nodes metrics – higher out-degree

In the third moment of analysis, as we can see by **Table 21**, the Ego has established new connections and received some new ones too, which made his value of *closeness degree* increase since the number of nodes did not increase as much as the new connections he has established. The graph has at this moment 354 nodes and 10.868 connections.

ID	In-degree	Out-degree	Closeness	Betweenness
0003	213	214	0,723	0,261
1158	27	170	0,662	0,016
2684	4	170	0,669	0,008
2703	1	146	0,635	0,004
2998	21	136	0,612	0,019
0308	93	123	0,570	0,014

Table 21 - Moment 3 - Top nodes metrics – higher out-degree

Summing up, we can verify that over the three years, the Ego remained the node of its own Ego Network with the highest value.

3.3.4 **0003 Ego Network (without the Ego)**

3.3.4.1 Network-level analysis

As we have done in the previous topics, **Figure 30** presents 0003's Ego Network without the presence of node 0003 itself, according to the depth of the network. Using *Gephi*'s capabilities, we are also able to obtain node 0103's Ego Network excluding the node's own presence and, consequently, its connections, which results in the graphs represented in **Figure 23**.

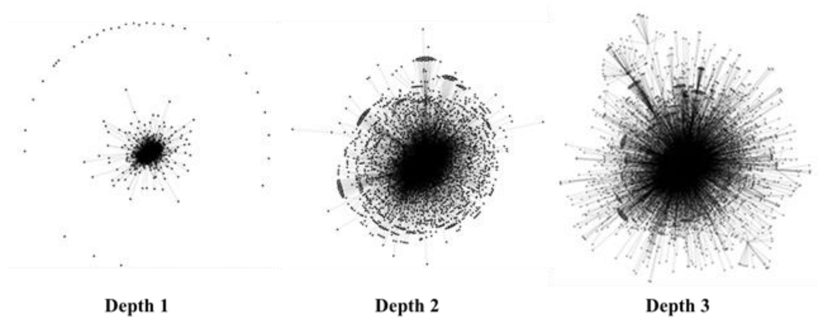

Figure 30 - Node 0003 trust network without node 0003

Table 22 represents the main network-level metrics of each of the graphs, which, as we have previously done, we designated Graph D followed by ".1", ".2" or ".3", according to the depth of the network.

Metrics	Graph D.1	Graph D.2	Graph D.3
Nodes	371	2880	4834
Edges	6164	25926	28617
Average degree	15,26	8,905	5,884
Density	0,038	0,003	0,001
Network diameter	7	8	10
Modularity	0,11	0,174	0,253
Connected components	33	29	28
Avg. Path length	2,611	3,46	3,93

Table 22 - Node 0003's Ego Network without the ego metrics

If we compare the number of nodes from Graph D.1 with the number of nodes from Graph C.1, we can see that the difference results almost entirely from the disappearance of Node 0003's direct connections, whose out-degree equals 404, as we have seen in **Table 16**.

3.3.4.2 Node-level analysis

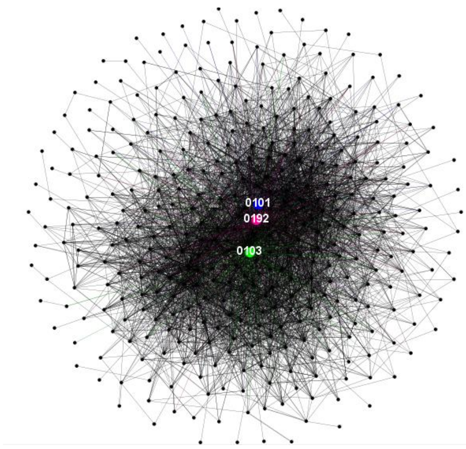

Figure 31 - Graph D.1 Top nodes

Once again, as the Ego has been excluded, other nodes will stand out.

Figure 31 is a representation of Graph D.1 top nodes. In blue, we have node 0101 which, as we will verify, presents the highest value of the *closeness degree;* in pink, there is node 0192, which has the highest value of *betweenness* and the second highest value of *closeness*; finally, in green, there is node 0103, which has the second highest value of *betweenness*.

ID	In-degree	Out-degree	Closeness	Betweenness
0101	19	134	0,594	0,034
0192	74	131	0,593	0,063
0676	46	116	0,566	0,025
1493	44	101	0,546	0,022
0092	18	80	0,537	0,011
0443	3	80	0,533	0,005
0595	23	71	0,528	0,018
0017	65	77	0,525	0,030
0028	45	77	0,523	0,020
0137	65	65	0,507	0,025

Table 23 - Graph D.1 Top closeness

Regarding Graph D.1, if we compare **Table 23** with **Table 16**, we can see that, beside the disappearance of the Ego, there are some nodes that stand out significantly to the detriment of other nodes. Just by looking at both tables, we can notice that nodes 0073 and 0152 are no longer in the Top 10 nodes with higher values of *closeness* and, for example, node 0192 has a significant growth on its *closeness* value, as well as node 0676, as we can see in **Table 24**.

ID	Closeness C1	Closeness D1	Var. Closeness
0101	0,600	0,594	-1%
0192	0,525	0,593	13%
0676	0,509	0,566	11%
1493	0,495	0,546	10%
0092	0,556	0,537	-3%
0443	0,486	0,533	10%
0595	0,483	0,528	9%
0017	0,481	0,525	9%
0028	0,481	0,523	9%
0137	0,468	0,507	8%

Table 24 - Graph C.1 *vs.* Graph D.1 Top closeness

In terms of *betweenness*, with the absence of the Ego, there was no place for a very significant highlight of any node. There were a few nodes who gained some position as intermediaries but still with reduced values, as we can see in **Table 25**.

ID	Betweenness C1	Betweenness D1	Var. Betweenness
0192	0,061	0,063	3%
0103	0,056	0,056	0%
0101	0,038	0,034	-12%
0017	0,031	0,030	-3%
0137	0,026	0,025	-4%
0676	0,023	0,025	4%
1493	0,019	0,022	12%
0090	0,020	0,021	4%
0028	0,018	0,020	10%
0595	0,013	0,018	37%

Table 25 - Graph C.1 *vs.* Graph D.1 top betweenness

In Graph D.2, with the increase of the number of edges, it is natural that both the values of *closeness* and *betweenness* have a decrease, as we have also seen comparing the metrics from Graph C.1 to Graph C.2, and that the nodes in the top 10 do not change a lot.

In fact, regarding the *closeness degree,* as we can verify in **Table 26**, the top nodes do not change from Graph D.1 to Graph D.2. The only thing that changes is the number of nodes they connect to and, consequently, its proximity to all the other nodes.

ID	Closeness D1	Closeness D2	Var. Closeness
0101	0,594	0,473	-20%
0192	0,593	0,473	-20%
0676	0,566	0,452	-20%
1493	0,546	0,447	-18%
0017	0,525	0,442	-16%
0595	0,528	0,438	-17%

Table 26 - Graph D.1 *vs.* Graph 2 Top closeness

In Graph D.3, the nodes we have highlighted in **Figure 27** are still the nodes with the highest values of *betweenness* and *closeness*, with node 0192 being the one with the higher values on both metrics, as we can see in **Table A. 11** and **Table A. 12**.

3.3.4.3 Dynamic analysis

With this analysis, we intend to understand how the network behaves, considering that we will exclude the Ego from the analysis, which was the node the highest out-degree of the network. We believe that with the exclusion of the Ego, the values of *closeness* of the remaining nodes will decrease, since to reach the same node they will now have to go through a longer path.

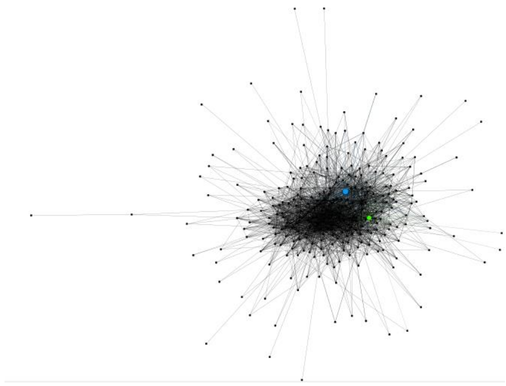

Figure 32 - Moment 1 - without the Ego

In the first period of analysis – Jan 1st, 2001 to Feb 28th, 2001, the graph has 289 nodes and 4106 edges, which means that with the absence of the Ego there were 16 nodes that were only connected to the Ego.

ID	In-degree	Out-degree	Closeness	Betweenness
2998	12	101	0,587	0,038
8440	5	90	0,562	0,009
8866	8	83	0,559	0,014
0076	48	79	0,518	0,047
1158	8	75	0,538	0,013

Table 27 - Moment 1 w/o the Ego - Top nodes metrics - higher out-degree

Table 28 shows that, as we were expecting, the *closeness degree* decreases for every node in **Table 27**, comparing to **Table 19**.

ID	Closeness_w/ Ego	Closeness_w/o Ego	Var.
2998	0,594	0,587	-1,24%
8440	0,572	0,562	-1,76%
8866	0,576	0,559	-2,87%
0076	0,529	0,518	-2,01%
1158	0,526	0,538	2,31%

Table 28 - Comparison between Moment 1 with and without the Ego

In the second moment of analysis, the network had 327 nodes and 8142 edges, which means that some of the nodes that were new to the network at moment 2, when we were considering the Ego, were only connected to him.

ID	In-degree	Out-degree	Closeness	Betweenness
1158	27	169	0,667	0,028
2684	4	169	0,665	0,009
2703	1	145	0,635	0,004
2998	21	135	0,608	0,031
0076	68	98	0,545	0,046

Table 29 - Moment 2 w/o the Ego - Top nodes metrics - higher out-degree

As shown in **Table 30**, when we compare the *closeness degree* of this moment to the same moment but including the Ego in the analysis, we see that it isn't as high.

ID	Closeness w/ Ego	Closeness_w/o Ego	Var
1158	0,674	0,667	-1,09%
2684	0,667	0,665	-0,27%
2703	0,639	0,635	-0,56%
2998	0,611	0,608	-0,46%
0076	0,553	0,545	-1,48%

Table 30 - Comparison between Moment 2 with and without the Ego

In the third moment, the graph has 335 nodes and 10.441 edges and node 1158 is still the node with the highest out-degree, as we can confirm by **Table 31**.

ID	In-degree	Out-degree	Closeness	Betweenness
1158	27	169	0,662	0,024
2684	4	169	0,662	0,009
2703	1	145	0,633	0,004
2998	21	135	0,608	0,027
0308	92	122	0,556	0,017
0076	79	118	0,570	0,046

Table 31 - Moment 3 w/o the Ego - Top nodes metrics - higher out-degree

3.3.5 **Summary**

In this sub-topic will be made a brief summary of the analysis we did in sections 3.3.1, 3.3.2, 3.3.3 and 3.3.4, as well as a contextualization to our trust network.

We started chapters 3.3.1 and 3.3.3 by presenting nodes 0103 and 0003's trust Ego Networks, which were the nodes we had previously identified as having the highest values of *in-degree* and *out-degree.*

Specifically, node 0103 and node 0003 are users of *Epinions.com* to whom, for simplicity, an identification number has been assigned.

For each node, we present three different graphs, depending on the depth of the network. When equal to 1, it represents the people in whom user 0103, for instance, trusts and whose reviews he follows; when equal to 2, represents the people from depth 1 plus the people in which the people user 0103 trusts trust, which are the people that will appear in user 0103's recommendations and that may sometimes appear during its searches, so these people are not totally unknown to node 0103 – are their acquaintances. When the depth is 3, appear all the people from depth 2 network plus the friends of node 0103's acquaintances.

With this in mind, it is natural that, by increasing the depth of the network, the number of users directly or indirectly connected to both user 0103 or 0003 increases as well.

When we mentioned that user 0103 had the highest value of *in-degree* it meant that he is the user other users trust the most. In this sense, being this user more popular than influential – people trust him more than he trusts in people and, therefore, their trust connections do not have as much influence on other people's trust connections – the more the network depth is increased, the less central is his position in his own network – his metrics values decrease in terms of *closeness* and *betweenness*.

On the other hand, being user 0003 the one with the highest value of *out-degree*, it means that this user trusts more people than people trust him. In this context, we can say that user 0003 is an influential user, since people that trust him start seeing in their recommendations more users that, supposedly, will have similar interests as themselves, according to the principles of homophily and transitivity.

In chapter 3.3.2 and 3.3.4, we focused on the same networks but excluded the Ego's presence which means that the users who, from the whole network, only trusted the Ego will be isolated and, therefore, no longer part of it.

With this analysis, we were able to see what impact the exclusion of the Ego from its own Ego Network has on the structure of the network itself and on the importance of the other nodes, since what happens, in fact, is that the trust network formed by each one of the Egos – whether it was user 0103 or 0003 – we exclude his own presence, leaving only not only the users on whom the Ego trusts and / or who rely on the Ego but also who also has trust connections between them, since with the exclusion of the Ego, all users who were only linked to the Ego itself, are also excluded from the network.

3.4 Communities evolution

With the aim of detecting communities and their evolution throughout the three years, we assessed the modularity value of the network, with depth 1, in four different moments, and we have done it in the network, firstly, including and secondly excluding the Ego.

We will analyze node 0103's ego network from January 1st, 2001 to August 31st, 2003, whose network has a total of 632 nodes and 14.911 edges.

Our first period of analysis is between Jan 1st, 2001 and Feb 28th, 2001 and, at this moment, the network has 556 nodes and 9.324 edges.

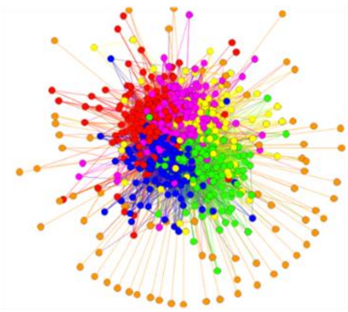

Figure 33 – Node 0103 Ego Network - Communities - period 1

In **Figure 33** we can distinguish 6 main components by its colors: red, which is formed by 17,09% of the nodes, pink (16,46%), blue (15,51%), green (15,03%), yellow (12,34%) and amber (11,55%). This means that these 6 components are formed by 87,98% of the nodes of the network. Each component is designated as community.

In **Table 32** are presented some metrics about the network.

Metric	Value
Average degree	14,753
Graph Density	0,023
Modularity	0,162

Table 32 - Node 0103 Ego Network - Metrics - period 1

The second period is from Jan 1st, 2001 to Feb 28th, 2002, which means that it now considers another year. In that period time, the network had 65 new nodes and 4.376 new trust connections were made, which means that the network has now 621 nodes and 13.700 edges.

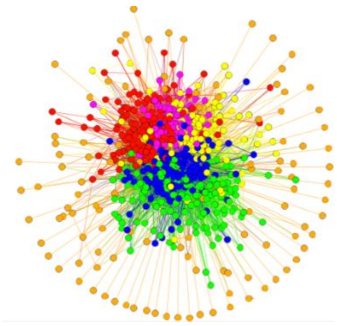

Figure 34 – Node 0103 Ego Network - Communities - period 2

As we can see, in this period, the network kept its 6 main communities but this time, as shown in **Table 33**, the red community is now formed by 18,35% of the nodes; the pink community lost some nodes from period 1 to period 2 that mainly joined the red and yellow communities, in fact, 32 nodes from the pink community joined the red community and 12 nodes joined the yellow community, which means that the pink community is now formed by 9,65% of the nodes; the blue community has 18,2% of the nodes; the green has 20,41%; the yellow has 13,92% and, finally, the amber community has 17,72%.

Community	Period 1	Period 2
Red	17,09%	18,35%
Pink	16,46%	9,65%
Blue	15,51%	18,20%
Green	15,03%	20,41%
Yellow	12,34%	13,92%
Amber	11,55%	17,72%

Table 33 - Constitution of the communities (% nodes)

As the number of edges had an increase higher than the increase of the number of nodes (46,93% > 11,69%), it is expected that the density of the graph in period 2 is higher than it was in period 1 and it is also expected that the value of the average degree increases as well, and we can confirm it by looking at **Table 34**.

Metric	Period 1	Period 2
Average degree	14,753	21,677
Graph Density	0,023	0,034
Modularity	0,162	0,149

Table 34 - Node 0103 Ego Network - Metrics - period 1 and 2

The third period is until Feb 28th, 2003 and at this moment the network has 628 nodes and 14.698 edges which means that in terms of the density of the network it is not expected that is changes much, since the increase in the number of nodes was very residual and the increase in the number of edges was only of 7,28%, which will make the density value to increase slightly, as we can see in **Table 35**.

Metric	Period 1	Period 2	Period 3
Average degree	14,753	21,677	23,256
Graph Density	0,023	0,034	0,037
Modularity	0,162	0,149	0,147

Table 35 - Node 0103 Ego Network - Metrics - period 1 to 3

Regarding the structure of the communities, the network continues with the 6 main communities but, just by looking at **Figure 35**, we can verify that some communities have lost some of their members, such as the blue community, that is now much smaller in favor of the green community that got much bigger.

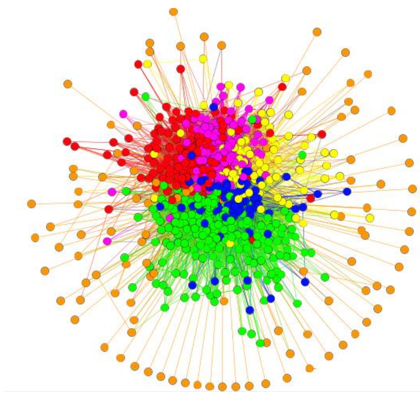

Figure 35 - Node 0103 Ego Network - Communities - period 3

As shown in **Table 36**, the green community had a great growth, followed by the pink community, to the detriment, mainly, of the blue community.

Community	Period 1	Period 2	Period 3
Red	17,09%	18,35%	18,67%
Pink	16,46%	9,65%	14,72%
Blue	15,51%	18,20%	10,28%
Green	15,03%	20,41%	28,16%
Yellow	12,34%	13,92%	11,39%
Amber	11,55%	17,72%	16,14%

Table 36 - Constitution of the communities (% nodes)

The final period of analysis is the whole period, that is, from Jan 1st, 2001 to Aug 31st, 2003. As we've said, the network has 632 nodes and 14.911 edges. Once again, the increase in the number of nodes is not that big, neither is the increase in the number of edges, which allows us to deduce that the metrics of the network will not change significantly.

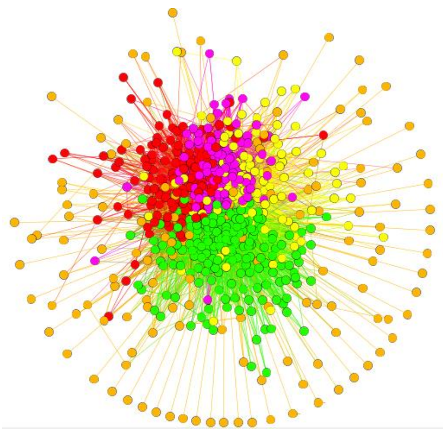

Figure 36 - Node 0103 Ego Network - Communities - period 4

Regarding the communities of the network, we can see by **Figure 36** and **Table 37** that the blue community has completely disappeared.

Community	Period 1	Period 2	Period 3	Period 4
Red	17,09%	18,35%	18,67%	19,30%
Pink	16,46%	9,65%	14,72%	15,98%
Blue	15,51%	18,20%	10,28%	-
Green	15,03%	20,41%	28,16%	28,48%
Yellow	12,34%	13,92%	11,39%	15,03%
Amber	11,55%	17,72%	16,14%	21,20%

Table 37 - Constitution of the communities (% nodes)

As we were expecting, due to the small growth in both the number of nodes and connections, the graph density did not change and the average degree has only increased slightly.

Metric	Period 1	Period 2	Period 3	Period 4
Average degree	14,753	21,677	23,256	23,593
Graph Density	0,023	0,034	0,037	0,037
Modularity	0,162	0,149	0,147	0,141

Table 38 - Node 0103 Ego Network - Metrics - period 1 to 4

Comparing the variation of the metrics in the four periods of analysis - **Table 38** – we realized that the periods were all very similar, this is, there were no major changes over time; however, between period 1 and 2 it is noticed that there was a superior growth both in the number of users of the website as well as in the number of trust connections.

Overall, we can say that, over time, user 0103's Ego Network has become denser and that users – his contacts – have increased the trust connections between themselves.

Given the analysis that we have just done, we will now repeat it but, this time, we will not include the Ego in the network, this is, we will analyze user 0103's Ego Network but disregarding its own presence as well as all the connections that were only connected to him.

Therefore, once again, we assessed the modularity of the network at each of the already defined periods of analysis and analyze its metrics and communities.

In the first period of analysis, between Jan 1st, 2001 and Feb 28th, 2001 the network has 523 nodes and 8.758 edges, which means that the Ego had 32 nodes that were only connected to him that were now excluded from the network. This will impact not only the communities of the network but, consequently, its metrics.

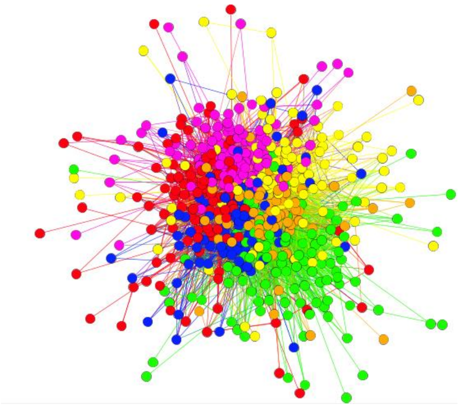

Figure 37 - Node 0103 Ego Network w/o Ego - Communities - period 1

Figure 37 allows us to see that there are 6 main communities in this network, as it happened when we were considering the Ego. **Table 39** gives us the percentage of nodes that form each of the communities. The red and the green communities are the biggest ones and the amber community is the smaller.

Community	Period 1
Red	15,98%
Pink	15,19%
Blue	11,87%
Green	15,98%
Yellow	13,77%
Amber	9,97%

Table 39 - Constitution of the communities (% nodes)

When we were considering the network with the Ego, the red community was also the biggest one and the amber one was the smallest.

Regarding the metrics of the network, since both the number of nodes and the number of edges decrease, facing the network with the Ego, almost in the same proportion (number of nodes has decreased in 6,3% and the number of edges 6,07%), it is expected that the value of the graph's density does not suffer a substantial change; nevertheless, since the number of nodes has decreased more that the number of edges, the density of the network should increase as well as the average degree, which we can confirm by **Table 40**.

Metric	Period 1
Average degree	16,746
Graph Density	0,032
Modularity	0,166

Table 40 - Node 0103 Ego Network w/o Ego - Metrics - period 1

In the second period of analysis, the network has 577 nodes – 54 new nodes were added to the network in the meanwhile, facing the 63 new nodes when we were considering the Ego – and 12.947 edges.

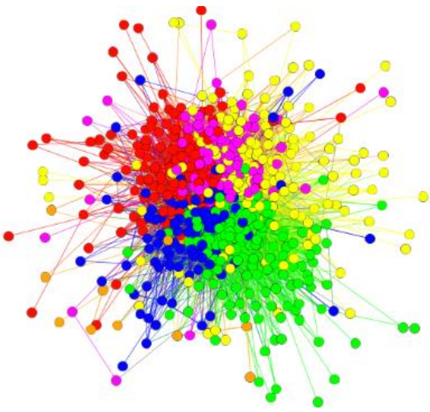

Figure 38 - Node 0103 Ego Network w/o Ego - Communities - period 2

At this moment, the network has kept its 6 main communities but the amber one has had a significant loss of the nodes that constituted it.

Community	Period 1	Period 2
Red	15,98%	19,30%
Pink	15,19%	12,82%
Blue	11,87%	14,08%
Green	15,98%	21,84%
Yellow	13,77%	21,36%
Amber	9,97%	1,74%

Table 41 - Constitution of the communities (% nodes)

As we can see on **Table 41**, the two major communities are the green and yellow ones. When we were considering the Ego, the two biggest communities were the green and the red ones. In fact, the yellow community was the second smaller, as we may verify in **Table 34**. This means that, without the Ego, some of the nodes that were in the orange community joined the yellow community.

Since the number of edges has had a significant growth (47,83%), compared to the increase of the number of nodes (10,11%), the density of the network and the average degree will increase as well, as presented in **Table 42**.

Metric	Period 1	Period 2
Average degree	16,746	22,438
Graph Density	0,032	0,039
Modularity	0,166	0,156

Table 42 - Node 0103 Ego Network w/o Ego - Metrics - period 1 and 2

Just like what happened when the Ego was being considered, from period 2 to period 3 the growth in the number of nodes was very low, compared to the growth from period 1 to period 2. The network, in the third period, has 585 nodes and 13.914 edges, which means that there are more new connections than users and, therefore, the network will be denser, as we can see in **Table 43**.

Metric	Period 1	Period 2	Period 3
Average degree	16,746	22,438	23,785
Graph Density	0,032	0,039	0,041
Modularity	0,166	0,156	0,146

Table 43 - Node 0103 Ego Network w/o Ego - Metrics - period 1 to 3

Regarding the community's structure, period 3 was of some changes. The amber community almost disappeared and the blue community has lost most of its nodes.

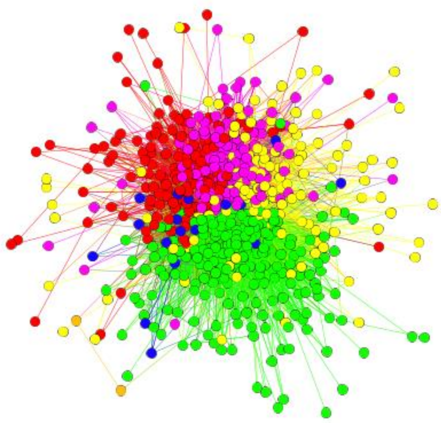

Figure 39 - Node 0103 Ego Network w/o Ego - Communities - period 3

As **Figure 39** suggests, the nodes from the blue community are now part of the green and red community. And we can confirm this by **Table 44**, where we see a large reduction in the percentage of nodes from the blue community and a significant increase in the percentage of nodes of the green community.

Community	Period 1	Period 2	Period 3
Red	15,98%	19,30%	20,41%
Pink	15,19%	12,82%	17,72%
Blue	11,87%	14,08%	2,85%
Green	15,98%	21,84%	32,28%
Yellow	13,77%	21,36%	18,99%
Amber	9,97%	1,74%	0,32%

Table 44 - Constitution of the communities (% nodes)

In the fourth period, the network has 590 nodes and 14.119 edges. That means that the absence of the Ego caused 42 nodes to no longer be part of this network.

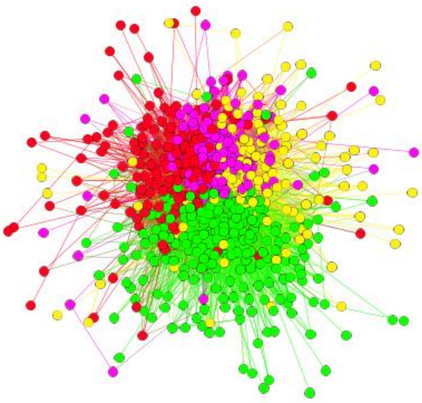

Figure 40 - Node 0103 Ego Network w/o Ego - Communities - period 4

In this last period of analysis, both blue and amber communities have disappeared, as we can see by **Figure 40** and **Table 45**.

Community	Period 1	Period 2	Period 3	Period 4
Red	15,98%	19,30%	20,41%	23,42%
Pink	15,19%	12,82%	17,72%	16,77%
Blue	11,87%	14,08%	2,85%	-
Green	15,98%	21,84%	32,28%	35,28%
Yellow	13,77%	21,36%	18,99%	17,72%
Amber	9,97%	1,74%	0,32%	-

Table 45 - Constitution of the communities (% nodes)

Compared to **Table 38**, we can see that, with the absence of the Ego, the network lost the amber community.

Just to finish this topic, it is important to emphasize that the methodology of calculating the modularity for each period of analysis has its limitations since the constitution of the communities was quite controlled, because we tried that the value of the modularity was not very different from one period to another and that the number of communities did not change a lot from period to period.

Since the modularity presents a random component, when we are assessing it in each period, it could reveal very different values from period to period, and that is the reason why we tried to control it, to avoid these differences.

Nevertheless, we considered it important to do this analysis to get an idea of how the network behaved over time in terms of the evolution of the communities, with and without the Ego.

4 Conclusions

Social Network Analysis is an increasingly useful practice, both online and offline. It allows us to know better and improve the organizational structure of an institution, for example, or even to predict what kind of behavior is more likely to occur in a certain situation, according to what's surrounding.

However, when we want to analyze a more particular case, for instance, the structure of a specific department in a company or the way an information flow circulated from a particular person, the most practical way is to resort to the analysis of Ego Networks, since they allow us to isolate a network from a specific entity and analyze it in greater depth.

In addition to being a type of analysis that is not yet as studied as the analysis of global networks, our intention was to perceive the impact, in the rest of the nodes of the network and in its own structure, that the exclusion of the Ego had it its own network.

With that goal in mind, it seemed to us that the analysis of a trust network would be very interesting, since trust is a characteristic that does not depend only on one entity – there must be necessarily two entities: the trustor and the trustee. This relationship of trust which, as we mentioned in chapter 2.2, does not necessarily have to be mutual, causes us to have a direct network.

In the specific case of this work, we analyzed a trust network extracted from *Epinions.com*. This network is an online trust-based network, where the users of the website have the option to create a list of people they trust based on the reviews of each person and where, according to the relationships of trust they establish, they can receive suggestions from people they may want to trust.

In order to carry out this work, we began by making a brief analysis of the global network, where we identified two users – node 0103 and node 0003 – about whom we performed our analysis at the Ego level. We considered it interesting to carry out this analysis on these two users since one of them – node 0103 – is the user that most people trust and the other one – node 0003 – is the user who trusts more people, which meant that the networks could have different behaviors.

From our analysis, we were able to draw some conclusions:

- The networks have a start layout, as we can see from **Figure 41**;

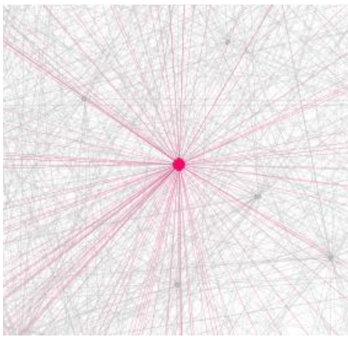

Figure 41 - Star-layout

- For our period of analysis, they were always sparse – the closer to 1 the density of graph is, the denser the network, and the highest density value our networks has was 0,041.

  As we increased the depth of the networks, the density value decreased, which means that the users who were added to the network did not have enough links to increase its density.

  On the other hand, with the dynamic analysis, we noticed that, over time, the density increased, once new nodes are added and the existing ones are establishing more links between them.

  When we compare the density values of the networks including and excluding the Ego, and with the different depths, we realize that there is no significant variation; in contrast, when, instead of changing the depth, we fix it as equal to 1, but we do an analysis over 3 years, we see that the exclusion of the Ego and the nodes that are only connected to him, makes the density increase, compared to the network in which we include the Ego.
- We noticed that the increase of the depth, including and excluding the Ego, does not have a significant impact that allows us to obtain some conclusion;
- As we have already mentioned, over the three years, the network has become a little denser but still very sparse. In terms of the evolution of the

communities, we could see that the modularity value of the networks does not indicate the existence of strong communities, since it is always well below 0,3. We saw that, as time went by, some communities declined – some even disappeared – and some, in turn, increased. This happened both in the network with the Ego and in the network without the Ego, however, when we excluded the Ego, the decline of some communities happened faster and, in the last period of analysis, has even lost two of the six communities, while when considering the Ego only one of the six communities disappeared. We believe that this happens because the Ego strengthens some relationships and when we exclude it, we make the connection of certain nodes to a group more weakened.

Summing up, we believe that this exploratory analysis helped us understand the extent to which the presence/absence of the Ego has an impact on its neighbors and its own network. We realize that, from a static perspective, in which we only vary the depth of the network, the absence of the Ego does not have a significant impact, the network develops approximately in the same way, nevertheless giving prominence to some nodes which, with the presence of the Ego, would not be so highlighted. From a dynamic perspective, we saw slight differences between networks – essentially because we are excluding from the analysis those nodes that were only connected to the Ego –, mainly in the evolution of communities, as already mentioned.

Contextualizing this summary with the reality of online trust networks, we believe that the little variation of the networks, when excluding the Ego from the analysis, means that, once a user establishes a relationship of trust with someone, the exclusion of other node you trust in – the Ego, is this case – will not have a major impact on the other user's relationships. As we have seen with the evolution of the communities, the exclusion of the Ego can cause one person to come closer to another but it does not imply any radical change.

The case could be different if this network was an offline network, since other variables would enter the equation of establishing or not a trust relationship that not only the quality of someone's reviews.

References

Abbott, K. M., Bettger, J. P., Hampton, K., & Kohler, H.-P. (2012). Exploring the use of social network analysis to measure social integration among older adults in assisted living. *Family & Community Health: The Journal of Health Promotion & Maintenance*, 322-333.

Adali, S., Escriva, R., Goldberg, M. K., Hayvanovych, M., Magdon-Ismail, M., Szymanski, B. K., . . . Williams, G. T. (2010). Measuring behavioral trust in social networks. *2010 IEEE International Conference on Intelligence and Security Informatics.* Vancouver, BC, Canada.

Arnaboldi, V., Conti, M., Passarella, A., & Pezzoni, F. (2012). *Analysis of Ego Network Structure in Online Social Networks.* Pisa, Italy.

Brooks, B., Hogan, B., Ellison, N., Lampe, C., & Vitak, J. (2014). *Assessing structural correlates to social capital in Facebook ego networks.*

Cambridge University Press. (2017, 05). *Cambridge online dictionary*. Retrieved from Cambridge Dictionary online: http://dictionary.cambridge.org/dictionary/english/trust

Davis, J., S. Chung, K. K., & Hossain, L. (2006). *Exploring Sociocentric and Egocentric Approaches for Social Network Analysis.* Australia: School of Information Technologies, University of Sidney.

Dunbar, R. (2010). *How many friends does one person need?* Cambridge, Massachusetts: Harvard University Press.

Dunbar, R. (2013, November 4). Robin Dunbar on Dunbar Numbers. (D. Edmonds, Interviewer) Social Science Bites. Retrieved from http://www.socialsciencespace.com/2013/11/robin-dunbar-on-dunbar-numbers/

Easley, D., & Kleinberg, J. (2010). Networks in their surrounding contexts. In D. Easley, & J. Kleinberg, *Networks, Crowds and Markets: Reaasoning about a Highly connected world* (pp. 85-118). Cambridge University Press.

Epinions.com. (2017, May). *Web archive*. Retrieved from Web archive: https://web.archive.org/web/20090405201239/http://www.epinions.com:80/help/faq/show_~faq_wot#010

Everett, M., & Borgatti, S. P. (2005). Ego network betweenness. *Social Networks*, 31-38.

Fortunato, S. (2010). Community detection in graphs. *Physics reports*.

Fu, B. (2007). *Trust Management in Online Social Networks.* University of Dublin, Trinity College.

Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., & Snasel, V. (2012). Social Network Analysis: Tools, Measures and Visualization. In A. Abraham, *Computational Social Network: Mining and Visualization* (pp. 3-23). Springer-Verlag London.

Golbeck, J., Parsia, B., & Hendler, J. (203). Trust networks on the semantic web. *Cooperative information agents VII*, 238-249.

Granovetter, M. (1983). The Strength of Weak Ties: a network theory revisited. In M. Granovetter, *Sociological Theory* (pp. 201-233). Wiley.

Guha, R., Kumar, R., Raghavan, P., & Tomkins, A. (2004). Propagation of trust and distrust. *Proceedings of the 13th international conference on World Wide Web* , (pp. 403-412). New York, USA.

Halgin, D. S., & Borgatti, S. P. (2012). *An Introduction to Personal Network Analysis and Tie Churn Statistics using E-NET.* Connections.

Hanneman, R. A., & Riddle, M. (2005). *University of California, Riverside.* Retrieved from University of California, Riverside - UCR Computing & Communications: http://faculty.ucr.edu/~hanneman/nettext/C7_Connection.html#geodesic

Hawley, K. (2014). *Trust, Distrust and Commitment.* Noûs, 48:1 1-20 DOI: 10.1111/nous.12000.

Hirst, A. J. (2010, May 10). *Ouseful.ifo*. Retrieved from https://blog.ouseful.info/: https://blog.ouseful.info/2010/05/10/getting-started-with-gephi-networkvisualisation-app-%E2%80%93-my-facebook-network-part-iii-ego-filters-and-simple-network-stats/

Josang, A., Hayward, R., & Pope, S. (2006). Trust network analysis with subjective logic. *Proceedings of the 29th Australasian Computer Science Conference-Volume 48* (pp. 85-94). Australian Computer Society, Inc.

Koçak, G. N. (2014). Social Networks and Social Network Analysis. *International Journal of Business and Social Science*, 5 (2), 126-135.

Kuz, A., & Giandini, R. (2012). *Social Network Analysis: a practical measurement and evaluation of Trust in a classroom environment.* XVIII Congreso Argentino de Ciencias de la Computación. Retrieved from http://hdl.handle.net/10915/23796

Leskovec, J., & Krevl, A. (2017, May). *SNAP Datasets: Stanford Large Network Dataset Collection.* Retrieved from https://snap.stanford.edu/data/soc-Epinions1.html

Massa, P., & Avesani, P. (2007). Trust-aware recommender systems. *Proceedings of the 2007 ACM conference on Recommender systems*, (pp. 17-24). Minneapolis, USA.

McAuley, J., & Leskovec, J. (2013). *Discovering Social Circles in Ego Networks.*

Moreno, J. L. (1934). *Who Shall Survive.* Washington D.C.: Nervous and Mental Disease Publishing Co.

Newman, M. E. (2004). Detecting community structure in networks. *The European Physical Journal B*, 321–330.

Newman, M. E. (2006). Modularity and community structure in networks. In *Proceedings of the National Academy of Sciences* (pp. 8577-8582). USA.

Oliveira, M., & Gama, J. (2012). An overview of Social Network Analysis. *Wiley interdisciplinary Reviews Data Mining and Knowledge Discovery, 2. 2,* pp. 99-15. John Wiley & Sons, Inc.

Oxford University Press. (2017, May). *Oxford Living Dictionaries*. Retrieved from OED Online: https://en.oxforddictionaries.com/definition/trust

Rad, P. A., Edzen, S., & Samuelsson, S. (2012). Measuring trust in online social networks: the effects of network parameters on the level of trust in trust games with incomplete information. *Proceedings of the 8th International Conference on Web Information Systems and Technologies* (pp. 531-539). Porto, Portugal: Karl-Heinz Krempels; José Cordeiro, SciTePress.

Sisson, D. (2017, June). *Introduction to Ecommerce concepts: Trust & Trustworthiness*. Retrieved from Philosophe: http://philosophe.com/ecommerce/trust/

Snijders, T. A. (2012). *Transitivity and Triads.* Oxford.

Tilly, C. (2010). Cities, states, and trust networks: chapter 1 of Cities and States in World History. In M. Hanagan, & C. Tilly, *Theory and Society* (pp. 265-280). Springer Netherlands.

Turilli, M., Vaccaro, A., & Taddeo, M. (2010). *The case of on-line trust.* Know Tech Pol 23:333 DOI: 10.1007/s12130-010-9117-5.

Victor, P., Cornelis, C., & De Cock, M. (2006). *Trust networks in recommender systems.* Ghent University.

Volakis, N. (2011). *Trust in Online Social Networks.* University of Edinburgh.

Wasserman, S., & Faust, K. (1994). Centrality and Prestige. In S. Wasserman, & K. Faust, *Social Network Analysis: Methods and Applications* (pp. 167-220). Cambridge: Cambridge University Press.

Wielens, J. (2014). *Ego Network Analysis: An Overview.* Mannheim.

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social Media Mining: An introduction.* Cambridge: Cambridge University Press.

Zolfaghar, K., & Aghaie, A. (2011). *Evolution of trust networks in social web applications using supervised learning.* Procedia Computer Science.

Appendix

id	indegree	outdegree	degree	closnesscentrality	betweenesscentrality
0103	265	199	464	0,433	0,071
0192	155	288	443	0,467	0,036
0003	15	404	419	0,510	0,034
0298	45	211	256	0,444	0,030
0017	140	172	312	0,443	0,027
0101	24	248	272	0,478	0,022
1493	109	243	352	0,450	0,022
0015	152	224	376	0,422	0,021
1977	9	71	80	0,354	0,018
0028	88	146	234	0,423	0,016

Table A. 1 - Top 10 betweenness - Graph A.2

ID	In-degree	Out-degree	Closeness	Betweenness
1469	2	1	1	0,00
5119	7	1	1	0,00
1946	4	1	1	0,00
4335	28	1	1	0,00
1560	7	1	1	0,00
4316	26	1	1	0,00
2017	3	1	1	0,00
1234	0	1	1	0,00
2277	1	1	1	0,00
2399	0	1	1	0,00
4545	4	1	1	0,00
3495	0	1	1	0,00

Table A. 2 - Graph A.2 closeness equal to 1

id	indegree	outdegree	degree	closnesscentrality	betweenesscentrality
0003	15	404	419	0,510	0,034
0101	24	248	272	0,478	0,022
0192	155	288	443	0,467	0,036
1493	109	243	352	0,450	0,022
0676	77	228	305	0,447	0,010
0298	45	211	256	0,444	0,030
0017	140	172	312	0,443	0,027
0196	74	73	147	0,443	0,016
0092	24	131	155	0,443	0,006
0152	40	120	160	0,435	0,005

Table A. 3 - Top 10 closeness - Graph A.2

ID	Betweenness_B1	Betweenness_B2	Var. Betweenness
0017	0,068	0,028	-59%
0147	0,042	0,01	-76%
1493	0,04	0,024	-40%
0196	0,04	0,016	-60%
0032	0,04	0,01	-75%
0201	0,036	0,011	-69%
0298	0,032	0,031	-3%
0714	0,032	0,008	-75%
0028	0,03	0,017	-43%
0010	0,028	0,016	-43%

Table A. 4 - Graph B.1 Top 10 betweenness vs. values from Graph B.2

ID	Betweenness_B2
0192	0,041
0003	0,036
0298	0,031
0017	0,028
0101	0,025
1493	0,024
0015	0,024
1977	0,018
0028	0,017
0231	0,017

Table A. 5 - Graph B.2 Top 10 betweenness

ID	Betweenness_A3	Betweenness_B3	Var. Betweenness
0192	0,030	0,031	5%
0101	0,022	0,023	3%
0015	0,020	0,022	8%
1493	0,018	0,019	3%
0003	0,017	0,018	7%
0017	0,017	0,018	4%
0298	0,015	0,015	1%
0204	0,013	0,013	2%
0178	0,011	0,012	8%
0028	0,011	0,012	5%

Table A. 6 - Top 10 betweenness Graph A.3 vs. Graph B.3

ID	Closeness_A3	Closeness_B3	Var. Closeness
0003	0,457	0,457	-0,06%
0101	0,439	0,440	0,15%
0192	0,429	0,430	0,14%
0676	0,410	0,410	-0,11%
1493	0,409	0,408	-0,24%
0092	0,404	0,402	-0,49%
0017	0,404	0,402	-0,53%
0152	0,397	0,398	0,20%
0595	0,396	0,395	-0,21%
0196	0,395	0,394	-0,22%

Table A. 7 - Top 10 closeness Graph A.3 vs. Graph B.3

ID	In-degree	Out-degree	Closeness	Betweenness
0003	15	404	1,000	0,205
0192	75	131	0,525	0,061
0103	91	76	0,471	0,056
0231	36	38	0,525	0,047
0196	36	36	0,523	0,042
0101	20	135	0,600	0,038
0017	66	77	0,481	0,031
0137	66	65	0,468	0,026
0121	23	26	0,517	0,025
0676	47	116	0,509	0,023

Table A. 8 - Graph C.1 Top betweenness

ID	In-degree	Out-degree	Closeness	Betweenness
0003	15	404	0,455	0,017
0101	25	308	0,438	0,026
0192	161	316	0,427	0,030
0676	77	228	0,406	0,007
1493	109	243	0,405	0,017
0092	24	131	0,402	0,003
0017	140	172	0,399	0,016
0152	40	129	0,394	0,004
0595	33	171	0,393	0,011
0196	74	73	0,392	0,008

Table A. 9 - Graph C.3 Top closeness

ID	In-degree	Out-degree	Closeness	Betweenness
0092	161	316	0,427	0,030
0101	25	308	0,438	0,026
0103	265	199	0,374	0,026
0015	174	260	0,384	0,020
1493	109	243	0,405	0,017
0003	15	404	0,455	0,017
0017	140	172	0,399	0,016
0298	45	211	0,391	0,014
0204	156	128	0,365	0,012
0028	88	146	0,387	0,011

Table A. 10 - Graph C.3 Top betweenness

ID	In-degree	Out-degree	Closeness	Betweenness
192	160	316	0,430	0,030
101	24	307	0,430	0,026
676	76	228	0,409	0,007
1493	108	243	0,407	0,017
17	139	172	0,402	0,016
595	32	171	0,395	0,011
92	23	130	0,394	0,002
298	45	211	0,393	0,014
418	111	106	0,391	0,009
28	87	146	0,389	0,011

Table A. 11 - Graph D.3 Top closeness

ID	In-degree	Out-degree	Closeness	Betweenness
192	160	316	0,430	0,030
103	264	199	0,375	0,026
101	24	307	0,430	0,026
15	174	260	0,386	0,020
1493	108	243	0,407	0,017
17	139	172	0,402	0,016
298	45	211	0,393	0,014
204	155	128	0,366	0,012
28	87	146	0,389	0,011
595	32	171	0,395	0,011

Table A. 12 - Graph D.3 Top betweenness